



# Modélisation des signes dans les ontologies biomédicales pour l'aide au diagnostic.

Pierre Sidoine V. Donfack Guefack

## ► To cite this version:

Pierre Sidoine V. Donfack Guefack. Modélisation des signes dans les ontologies biomédicales pour l'aide au diagnostic.. Médecine humaine et pathologie. Université de Rennes, 2013. Français. NNT : 2013REN1B018 . tel-01057310

**HAL Id: tel-01057310**

**<https://theses.hal.science/tel-01057310>**

Submitted on 22 Aug 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Européenne de Bretagne*

pour le grade de  
**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**

*Mention : Génie biologique et médical*

**École doctorale Vie – Agro – Santé (VAS)**

présentée par

**Sidoine Pierre V. DONFACK GUEFACK**

Préparée à l'unité de recherche 936 (UMR936)  
Unité INSERM 936

Modélisation Conceptuelle des Connaissances Biomédicales

---

**Modélisation des  
signes dans les  
ontologies  
biomédicales pour  
l'aide au diagnostic**

**Thèse soutenue à la faculté de  
médecine de Rennes 1  
le 20 décembre 2013**

devant le jury composé de :

**M. Jean CHARLET**

Chargé de mission - AP-HP & INSERM / *rapporteur*

**M. Alain VENOT**

PU-PH - Université Paris Nord / *rapporteur*

**Mme. Anita BURGUN**

PU-PH - Université Paris Descartes / *examinateur*

**M. Bernard GIBAUD**

CR - Université Rennes 1 / *examinateur*

**Mme. Valérie BERTAUD-GOUNOT**

PU-PH - Université Rennes 1 / *co-directeur*

**M. Régis DUVAUFERRIER**

PU-PH - Université Rennes 1 / *directeur des travaux*

**M. Marc CUGGIA**

PU-PH - Université Rennes 1 / *examinateur*



# Résumé

**Introduction :** Établir un diagnostic médical fiable requiert l'identification de la maladie d'un patient sur la base de l'observation de ses signes et symptômes. Par ailleurs, les ontologies constituent un formalisme adéquat et performant de représentation des connaissances biomédicales. Cependant, les ontologies classiques ne permettent pas de représenter les connaissances liées au processus du diagnostic médical : connaissances probabilistes et connaissances imprécises et vagues.

**Matériel et méthodes :** Nous proposons des méthodes générales de représentation des connaissances afin de construire des ontologies adaptées au diagnostic médical. Ces méthodes permettent de représenter : (a) Les connaissances imprécises et vagues par la discrétisation des concepts (définition de plusieurs catégories distinctes à l'aide de valeurs seuils ou en représentant les différentes modalités possibles). (b) Les connaissances probabilistes (les sensibilités et les spécificités des signes pour les maladies, et les prévalences des maladies pour une population donnée) par la réification des relations ayant des arités supérieures à 2. (c) Les signes absents par des relations et (d) les connaissances liées au processus du diagnostic médical par des règles SWRL. Un moteur d'inférences abductif et probabiliste a été conçu et développé. Ces méthodes ont été testées à l'aide de dossiers patients réels.

**Résultats :** Ces méthodes ont été appliquées à trois domaines (les maladies plasmocytaires, les urgences odontologiques et les lésions traumatiques du genou) pour lesquels des modèles ontologiques ont été élaborés. L'évaluation a permis de mesurer un taux moyen de 89,34% de résultats corrects.

**Discussion-Conclusion :** Ces méthodes permettent d'avoir un modèle unique utilisable dans le cadre des raisonnements abductif et probabiliste, contrairement aux modèles proposés par : (a) Fenz qui n'intègre que le mode de raisonnement probabiliste et (b) García-crespo qui exprime les probabilités hors du modèle ontologique. L'utilisation d'un tel système nécessitera au préalable son intégration dans le système d'information hospitalier pour exploiter automatiquement les informations du dossier patient électronique. Cette intégration pourrait être facilitée par l'utilisation de l'ontologie du système.

**Mots-Clés :** Aide au diagnostic, Ontologie biomédicale, Modélisation des connaissances médicales, Sémiologie quantitative, Ontology design pattern, raisonnement abductif, connaissances de contrôle



# Abstract

**Introduction:** Making a reliable medical diagnosis requires the identification of the patient's disease based on the observation of signs. Moreover, ontologies provide an adequate and efficient formalism for medical knowledge representation. However, classical ontologies do not allow representing knowledge associated with medical reasoning such as probabilistic, imprecise, or vague knowledge.

**Material and methods:** In the current work, general knowledge representation methods are proposed. They aim at building ontologies fitting to medical diagnosis. They allow to represent: (a) imprecise or vague knowledge by discretizing concepts (definition of several distinct categories thanks to threshold values or by representing the various possible modalities), (b) probabilistic knowledge (sensitivity, specificity and prevalence) by reification of relations of arity greater than 2, (c) absent signs by relations and (d) medical reasoning and reasoning on the absent signs by SWRL rules. An abductive reasoning engine and a probabilistic reasoning engine were designed and implemented. The methods were evaluated by use of real patient records.

**Results:** These methods were applied to three domains (the plasma cell diseases, the dental emergencies and traumatic knee injuries) for which the ontological models were developed. The average rate of correct diagnosis was 89.34 %.

**Discussion-Conclusion:** In contrast with other methods proposed by Fenz and García-crespo, the proposed methods allow to have a unique model which can be used both for abductive and probabilistic reasoning. The use of such a system will require beforehand its integration in the hospital information system for the automatic exploitation of the electronic patient record. This integration might be made easier by the use of the ontology on which the system is based.

**Keywords:** Medical decision support systems – Biomedical ontology – Medical knowledge representation – Quantitative semiology – Ontology design pattern – Abductive reasoning – control knowledge



# Remerciement

Je tiens à remercier tout d'abord, mon directeur et ma co-directrice, le Pr Régis DUVAUFERRIER et le Pr Valérie BERTAUD-GOUNOT, pour leur indéfectible soutien et pour m'avoir fait confiance.

Je remercie très vivement les Professeurs Jean CHARLET et Alain VENOT pour avoir accepté d'être rapporteurs de ce document. Merci de l'intérêt que vous avez porté à mon travail et de vos remarques qui ont contribué à améliorer la qualité de ce manuscrit.

Je tiens à remercier les professeurs Anita BURGUN et Marc CUGGIA, et le docteur Bernard GIBAUD d'avoir accepté de participer au Jury de soutenance.

Mes remerciements vont particulièrement aux Docteurs Olivier DAMERON, Olivier DECAUX, Jérémy LASBLEIZ et Adrien BARTON, pour leur disponibilité, leurs suggestions et leurs encouragements.

Je tiens à remercier mon tuteur le Dr Pierre-Yves LEBAIL pour sa disponibilité.

L'ensemble de l'équipe à laquelle je suis fier d'appartenir a été d'un grand soutien. En particulier, je remercie Denis DELAMARRE.

J'exprime ma gratitude également aux stagiaires et internes pour leur aide, en particulier Faustine SERRAND-OBRY et Marion VALENCE.

J'exprime ma gratitude également à mes amis et frères pour leur aide dans la mise en place des solutions techniques pour les tests et les évaluations des méthodes, en particulier Alain Roland BAVOUA et Jean Martin CHEUMELEU NINKO.

Un grand merci à Judith ma petite sœur chérie pour ses multiples relectures qui ont contribué à améliorer la qualité des documents.

Je ne remercierai jamais assez mes parents pour m'avoir fait confiance, encouragé et soutenu durant tout ce temps.

Enfin, tout au fond de mon cœur, merci à Doris ma fiancée, pour sa patience, son amour, sa confiance et son soutien constant durant ces longues années. Sans elle, rien de tout cela n'aurait été possible.





# Sommaire

|  |    |
|--|----|
| <b>Chapitre 1</b>                                      | 19 |
| Introduction générale                                  | 19 |
| 1.1 Contextes et enjeux                                | 19 |
| 1.2 Objectifs de la recherche                          | 22 |
| 1.3 Plan du mémoire                                    | 23 |
| <b>Chapitre 2</b>                                      | 25 |
| Définitions  | 25 |
| 2.1 Le signe et le symptôme                            | 25 |
| 2.1.1 Le signe positif                                 | 25 |
| 2.1.2 Le signe négatif                                 | 26 |
| 2.2 La maladie et le syndrome                          | 26 |
| 2.3 Le diagnostic médical                              | 27 |
| 2.3.1 Le raisonnement déductif                         | 28 |
| 2.3.2 Le raisonnement inductif                         | 28 |
| 2.3.3 Le raisonnement abductif                         | 29 |
| 2.3.4 Le raisonnement par analogie                     | 29 |
| 2.4 La sémiologie quantitative                         | 29 |
| 2.4.1 Tableau de contingence                           | 30 |
| 2.4.2 Interprétations                                  | 32 |
| 2.5 Conclusion   | 32 |
| <b>Chapitre 3</b>                                      | 33 |
| État de l'art  | 33 |
| 3.1 La connaissance                                    | 33 |
| 3.2 La typologie des connaissances                     | 33 |
| 3.2.1 Les connaissances déclaratives                   | 33 |
| 3.2.2 Les connaissances procédurales                   | 33 |
| 3.2.3 Les connaissances de contrôle                    | 34 |
| 3.3 La représentation des connaissances                | 34 |
| 3.3.1 Les ontologies                                   | 35 |
| 3.3.2 Les langages de représentation des connaissances | 41 |

|                       |   |           |
|-----------------------|---|-----------|
| 3.3.2.1               | Les logiques de description .....   | 41        |
| 3.3.2.2               | Web Ontology Language : OWL.....  | 43        |
| 3.3.3                 | Système informatisé d'aide au diagnostic en urgences odontologiques .....                                     | 50        |
| 3.3.4                 | Ontology-driven differential diagnosis based on logical inference and probabilistic refinements (ODDIN) ..... | 51        |
| 3.4                   | Redéfinition des objectifs .....  | 62        |
| 3.4.1                 | Problématique .....   | 62        |
| 3.4.2                 | Objectifs .....   | 63        |
| 3.4.2.1               | Modèle de connaissance .....  | 63        |
| 3.4.2.2               | Système d'aide à la décision à base d'ontologie (ODES) .....  | 64        |
| 3.5                   | Conclusion .....  | 64        |
| <b>Chapitre 4</b>     | .....   | <b>65</b> |
| Matériels et méthodes | .....   | 65        |
| 4.1                   | Matériels .....   | 65        |
| 4.1.1                 | Le corpus .....   | 65        |
| 4.1.1.1               | L'ontologie du NCIT .....   | 65        |
| 4.1.2                 | Langage et outils de représentation des connaissances .....   | 72        |
| 4.1.3                 | Langages et outils de développement .....   | 73        |
| 4.1.3.1               | JESS .....  | 73        |
| 4.1.3.2               | Semantic Query Enhanced Web Rule Language: SQWRL .....  | 73        |
| 4.1.3.3               | Semantic Web Rule Language Interrogation Query : SWRL-IQ .....  | 74        |
| 4.2                   | Méthodes .....  | 74        |
| 4.2.1                 | L'opérationnalisation des modèles ontologiques .....  | 75        |
| 4.2.2                 | La représentation des signes absents ou négatif .....   | 83        |
| 4.2.3                 | Le raisonnement abductif .....  | 83        |
| 4.2.4                 | Le raisonnement probabiliste .....  | 85        |
| 4.2.5                 | La création des cas prototypiques .....   | 85        |
| 4.2.6                 | Les axiomes de fermeture (Closure axiom) .....  | 86        |
| 4.2.7                 | La création des ensembles .....   | 86        |
| 4.2.8                 | Le moteur d'inférences .....  | 87        |
| 4.2.9                 | Architecture et fonctionnement de l'Ontology Driven Expert System (ODES) .....                                | 87        |
| 4.2.10                | Évaluation .....  | 98        |
| 4.2.10.1              | Données médicales .....   | 98        |
| 4.2.10.2              | Méthodes d'évaluation .....   | 98        |

|                   |  |     |
|-------------------|--|-----|
| 4.3               | Conclusion .....   | 99  |
| <b>Chapitre 5</b> | .....  | 101 |
| Résultats         | .....  | 101 |
| 5.1               | Introduction .....   | 101 |
| 5.2               | Le modèle de connaissances .....   | 101 |
| 5.2.1             | La représentation dichotomique du signe.....                                   | 101 |
| 5.2.2             | La représentation multivaluée du signe.....                                    | 102 |
| 5.2.3             | La représentation de la probabilité d'un signe dans une maladie .....          | 103 |
| 5.2.4             | La représentation de la prévalence d'une maladie .....                         | 104 |
| 5.2.5             | La représentation des connaissances de contrôle ou de raisonnement.....        | 105 |
| 5.2.6             | La création des cas prototypiques .....  | 107 |
| 5.2.6.1           | La conservation de la hiérarchie des classes .....                             | 108 |
| 5.2.6.2           | La conservation de la profondeur des définitions.....                          | 108 |
| 5.2.6.3           | La base de règles d'ordre 0.....   | 109 |
| 5.2.7             | La représentation des axiomes de fermeture .....                               | 110 |
| 5.2.8             | Conclusion.....  | 110 |
| 5.3               | La description du système .....  | 110 |
| 5.3.1             | La création des ensembles et stockage dans une base de données relationnelle.. | 110 |
| 5.3.2             | Le moteur d'inférences mixte .....   | 111 |
| 5.3.3             | L'interface du système .....   | 112 |
| 5.3.3.1           | Le formulaire de connexion .....   | 112 |
| 5.3.3.2           | Le formulaire de choix de l'ontologie.....                                     | 113 |
| 5.3.3.3           | Le formulaire adaptatif.....   | 113 |
| 5.3.3.4           | Le formulaire de résultat .....  | 114 |
| 5.4               | Le calcul de la sémiologie quantitative .....                                  | 114 |
| 5.5               | Évaluation .....   | 117 |
| 5.5.1             | Matériels .....  | 117 |
| 5.5.2             | Méthodes.....  | 117 |
| 5.5.2.1           | Jeu de données 1 .....   | 117 |
| 5.5.2.2           | Jeu de données 2.....  | 117 |
| 5.5.2.3           | Jeu de données 3.....  | 117 |
| 5.5.3             | Résultats.....   | 118 |
| 5.6               | Conclusion.....  | 118 |

|  |     |
|--|-----|
| <b>Chapitre 6</b> .....  | 119 |
| Discussion et Conclusion .....   | 119 |
| 6.1 Introduction .....   | 119 |
| 6.2 Modèle de connaissances.....                                       | 119 |
| 6.2.1 Pourquoi une instanciation complète du modèle ontologique ?..... | 120 |
| 6.2.2 Le raisonnement abductif.....                                    | 121 |
| 6.2.3 Les signes absents .....   | 122 |
| 6.2.3.1 La représentation .....  | 122 |
| 6.2.3.2 Le raisonnement abductif et les signes absents .....           | 122 |
| 6.2.4 Ontologie et Probabilité .....                                   | 122 |
| 6.2.5 Les définitions multi-niveaux.....                               | 123 |
| 6.3 Système ODES .....   | 124 |
| 6.3.1 Évaluation.....  | 124 |
| 6.3.2 Temps de traitement .....  | 124 |
| 6.3.3 Calcul de la sémiologie quantitative .....                       | 124 |
| 6.4 Conclusion.....  | 125 |
| 6.5 Perspectives .....   | 126 |
| 6.5.1 A court terme .....  | 126 |
| 6.5.2 A moyen et long terme .....                                      | 126 |
| <b>Bibliographie</b> .....   | 127 |
| <b>Annexes</b> .....   | 143 |
| <b>Liste des publications et communications</b> .....                  | 153 |

# Tables des figures

|   |    |
|---|----|
| <b>Figure 1:</b> Structure de base d'un système d'aide à la décision. Des données sont fournies au système et le moteur d'inférences renvoie les résultats et les explications relatives à son raisonnement à partir du contenu de la base de connaissances. ....                             | 19 |
| <b>Figure 2:</b> Architecture des systèmes de représentation des connaissances s'appuyant sur les logiques de description (Baader 2003). ....   | 42 |
| <b>Figure 3 :</b> Écran de saisie d'une règle SWRL sous PROTÉGÉ 3.4.8. L'éditeur offre un système d'auto-complétion pour faciliter la saisie des règles. ....   | 49 |
| <b>Figure 4:</b> Description d'une maladie sous forme multi-niveau (DMN). ....  | 52 |
| <b>Figure 5:</b> Description d'une maladie dite simple. ....  | 52 |
| <b>Figure 6 :</b> Définition multi-niveau d'une maladie. « Disease Y » a une profondeur 2. ....   | 52 |
| <b>Figure 7:</b> Nouvelle définition de la classe Disease Y. La classe Disease X a été éclatée en mettant les signes de cette dernière (Sym C, Sym D, Sym E). ....  | 54 |
| <b>Figure 8:</b> Nouvelle définition de la classe Disease Y. ....   | 54 |
| <b>Figure 9 :</b> Une partie de la description d'une MGUS pour mettre en exergue les niveaux d'opérationnalisation d'une ontologie (Peelen et al. 2007). ....   | 57 |
| <b>Figure 10 :</b> Exemple de base de connaissances modélisée selon l'approche PCT (Henson et al. 2011). ....   | 61 |
| <b>Figure 11:</b> Une vue d'une partie de l'ontologie du NCIT dans l'éditeur Protégé 3.4.8. ....  | 66 |
| <b>Figure 12:</b> Une vue des relations de l'ontologie du NCIT dans Protégé 3.4.8. ....   | 67 |
| <b>Figure 13:</b> Une vue des relations de l'ontologie du NCIT dans Protégé 3.4.8 après la modification sur la hiérarchie des relations apportée par Natalya Noy et ses collègues (Noy et al. 2008). ....   | 67 |
| <b>Figure 14 :</b> L'évolution clinique du patient se fait parallèlement aux modifications histologiques (Sirohi and Powles 2004). ....   | 68 |
| <b>Figure 15:</b> Définition d'une MGUS dans le NCIT (Fragoso et al. 2004). On peut constater l'absence des critères CRAB, la définition très vague de la quantité de l'immunoglobuline ( <i>ligne sélectionnée</i> ). ....   | 69 |
| <b>Figure 16:</b> Définition d'une MGUS dans la littérature (Kyle et al. 2011; Rajkumar 2011). ....   | 70 |
| <b>Figure 17 :</b> Ontologie des signes cliniques des lésions traumatiques du genou. Cette version de l'ontologie ne représente que les connaissances qualitatives telles que : les lésions et les signes. Des catégories de lésions, de signes et de symptômes sont aussi représentées. .... | 71 |
| <b>Figure 18:</b> Spécialisation du concept IgA_Measurement en utilisant les données de la littérature. Les deux spécialisations sont respectivement des conditions de MM et MGUS. ..   | 77 |
| <b>Figure 19:</b> Représentation des connaissances bivaluées. Expression de la restriction : une MGUS a pour signe le «IgA_Measurement_Less_Than_3g_Per_dL». ....   | 77 |
| <b>Figure 20:</b> Expression de la restriction : une MGUS a pour signe le « IgA_Measurement < 3g/dL ». Cette restriction n'est pas exprimable directement en OWL car la relation « Disease_Has_Finding_With_ThresholdValue_and_Unit » est d'arité 4 ou d'arité n. ....                        | 78 |
| <b>Figure 21:</b> Expression des connaissances multivaluées. Nous avons réifié la relation d'arité 4 en créant une classe « MGUS_IgA_Measurement » représentant le couple (MGUS,IgA_Measurement). A cette nouvelle classe nous lions le signe IgA_Measurement, la                             |    |

valeur de la mesure du signe et l'unité. Toutes les nouvelles relations créées pour exprimer ces restrictions sont d'arité 2. Pour finir, nous lions la classe « MGUS » à notre classe « MGUS\_IgA\_Measurement ». ..... 79

**Figure 22:** Expression des connaissances probabilistes. Représentation de la restriction : une MGUS a une prévalence de 5% pour les patients de plus de 70 ans et a pour signe le « IgA\_Measurement < 3g/dL » avec une sensibilité de 25% et une spécificité de 90%. Trois relations d'arité 3 « Disease\_Prevalence\_Context, Disease\_Has\_Finding\_Se, Disease\_Has\_Finding\_Sp » sont nécessaires pour cette représentation. Ces trois relations ne sont pas exprimables comme telles par le langage OWL. Pour ce faire nous devons utiliser la réification comme ci-après à la Figure 23. .... 80

**Figure 23:** Expression des sensibilités et spécificités des signes pour les maladies. Représentation de la restriction : une MGUS a pour signe le « IgA\_Measurement\_Less\_Than\_3g\_Per\_dL » avec une sensibilité de 25% et une spécificité de 90%. .... 81

**Figure 24 :** Expression de la prévalence d'une maladie pour une population donnée. La population donnée dans notre exemple est celle des patients de plus de 70 ans et elle est matérialisée par la classe « Context ». Pour cela nous avons créé une classe représentant le couple (Maladie, Contexte) soit « MGUS\_Context ». Deux relations d'arité 2 dont une de type DatatypeProperty « Has\_Prevalence » et une de type ObjectProperty « Has\_Context » ont été ajoutées. .... 82

**Figure 25:** Définition du MM sous forme de maladie multi-niveau. Le MM est représenté ici sous forme de pathologie de niveau ou profondeur 2. L'hypercalcémie y est définie via la mesure du calcium sérique et l'anémie y est définie via la mesure de l'hémoglobine. .... 82

**Figure 26:** Diagramme des cas d'utilisation du système ODES. Les relations d'utilisation « include » stipulent que le cas d'utilisation source contient aussi le comportement décrit dans le cas d'utilisation destination. En d'autre terme pour réaliser l'objectif « sélection de l'ontologie », on utilise les objectifs « connexion ». Les relations d'extension « extend » stipulent que le cas d'utilisation source étend (précise) les objectifs ou le comportement du cas d'utilisation destination. En d'autres termes, l'objectif « Mise à jour de la base de connaissances - Apprentissage » est réalisé après « vérification des résultats ». .... 89

**Figure 27:** Vue du schéma général du système d'aide au diagnostic médical fondé sur une ontologie biomédicale. Sur ce schéma nous avons toutes les composantes du système, partant de la génération des tuples nécessaire pour l'aide au diagnostic à l'étape de mise à jour de ces tuples. Les étapes de construction de l'ontologie biomédicale ne sont pas mentionnées car elles sont faites en amont de la génération des tuples et de l'exploitation dudit système (Jovic et al. 2007). .... 91

**Figure 28:** Modèle conceptuel de la base de connaissances du système. Cette base de connaissances est utilisée pour stocker le contenu de l'ontologie sémiologique biomédicale. C'est une sorte de copie de l'ontologie, toutes les informations nécessaires pour la description des maladies y sont. Ce choix a été fait uniquement pour des raisons de performance et de facilité d'accès pour les tests et évaluations multicentriques. .... 93

**Figure 29:** Modèle conceptuel de la base des cas du système. Cette base des cas est utilisée pour stocker les informations sur les décisions prises par les médecins. Ces données sont ensuite utilisées pour l'évaluation des performances du système et aussi pour affiner le

|   |     |
|---|-----|
| modèle de connaissance via le processus d'apprentissage. Pour garantir l'intégrité et la cohérence de la base des cas, des entités types représentant les maladies, les signes et les relations ont été intégrées pour stocker le contenu de l'ontologie sémiologique biomédicale et ainsi conserver à ce niveau une partie de la sémantique du domaine. .... | 95  |
| <b>Figure 30:</b> Dans le cadre de la représentation des critères diagnostiques d'une MGUS, il est nécessaire de distinguer si la mesure de l'IgA est inférieure à 3g/dL ou non. Dans ce cas, la mesure de l'IgA peut être représentée par deux sous-classes représentant pour l'une une mesure d'IgA < 3g/dL et pour l'autre une mesure d'IgA ≥ 3g/dL. ....  | 101 |
| <b>Figure 31:</b> Définition formelle partielle d'une MGUS. ....  | 102 |
| <b>Figure 32:</b> Création de nouvelles relations de type « <i>ObjectProperty</i> ». ....   | 102 |
| <b>Figure 33:</b> Définition partielle de la classe MGUS_IgA_Measurement. ....  | 102 |
| <b>Figure 34:</b> Création de nouvelles relations de type « <i>DatatypeProperty</i> ». ....   | 102 |
| <b>Figure 35:</b> Définition formelle partielle de la classe MGUS_IgA_Measurement. ....   | 103 |
| <b>Figure 36:</b> Définition formelle partielle de la classe MGUS_IgA_Measurement_Less_Than_3g_Per_dL. ....   | 103 |
| <b>Figure 37:</b> Création de nouvelles relations de type DatatypeProperty pour la représentation des prévalences des maladies pour un contexte et, des sensibilités et des spécificités des signes pour les maladies. ....   | 103 |
| <b>Figure 38:</b> Définition formelle partielle de la classe MGUS_IgA_Measurement_Less_Than_3g_Per_dL. ....   | 103 |
| <b>Figure 39:</b> Création d'une nouvelle relation de type « <i>ObjectProperty</i> ». ....  | 104 |
| <b>Figure 40:</b> Définition formelle partielle de la classe MGUS_Context. ....   | 104 |
| <b>Figure 41:</b> Définition formelle de la classe MGUS_Context. ....   | 104 |
| <b>Figure 42:</b> Définition formelle partielle de l'hypercalcémie. ....  | 105 |
| <b>Figure 43:</b> Définition formelle partielle du myélome multiple. ....   | 105 |
| <b>Figure 44:</b> Définition formelle partielle d'une MGUS avec un signe absent ou négatif. ....  | 105 |
| <b>Figure 45:</b> Création de nouvelles relations de type « <i>ObjectProperty</i> » pour la mise en œuvre des règles SWRL. ....   | 105 |
| <b>Figure 46:</b> Expression d'une règle d'arité 2 pour des signes dichotomiques. Cet exemple concerne la règle 1. ....   | 106 |
| <b>Figure 47:</b> Règle d'arité 2 conçu via un chaînage des relations. Il permet aussi de raisonner lorsqu'on a des cas de définitions multi-niveaux des maladies. Cet exemple concerne la règle 9. ....  | 106 |
| <b>Figure 48:</b> Expression de deux règles d'arité n. Ces règles permettent de calculer les rapports de vraisemblance (positif et négatif) et de sélectionner les autres données nécessaires pour le moteur d'inférences probabiliste. Ce type de règles est exprimé via SWRL-IQ. Cet exemple concerne respectivement les règles 11 et 12. ....              | 106 |
| <b>Figure 49:</b> Définition formelle partielle d'une MGUS. ....  | 107 |
| <b>Figure 50:</b> Instance d'une MGUS. ....   | 107 |
| <b>Figure 51:</b> Instances des signes d'une MGUS. ....   | 107 |
| <b>Figure 52:</b> Hiérarchie des types de lésion osseuse. ....  | 108 |
| <b>Figure 53:</b> Illustration de la conservation de la hiérarchie des classes au niveau des instances. Toutes les instances des classes subsumées par la classe « Bone_Lesion » sont aussi liées à l'instance « MGUS_1 ». ....   | 108 |



|  |     |
|--|-----|
| <b>Figure 54:</b> Instances des signes du MM. La conservation de la profondeur des définitions des classes fait en sorte que la mesure du calcium sérique est liée directement au MM.....  | 108 |
| <b>Figure 55:</b> Illustration de la conservation de la profondeur des définitions des maladies. L'instance « MM_1 » est liée à l'instance « Hypercalcemia_1 » et à l'instance « SC_Measurement_More_Than_2_6mmol_Per_L_1 » .....  | 109 |
| <b>Figure 56:</b> Exemple de règles d'arité 2 pour le cas d'une gammapathie monoclonale à signification indéterminée.....  | 109 |
| <b>Figure 57:</b> Exemple de règles d'arité n pour le cas d'une gammapathie monoclonale à signification indéterminée.....  | 109 |
| <b>Figure 58:</b> Définition formelle partielle d'une MGUS avec un axiome de fermeture. Cette restriction signifie qu'une MGUS n'exclut que les lésions osseuses.....  | 110 |
| <b>Figure 59:</b> Logigramme du moteur d'inférences mettant en œuvre les modes de raisonnement abductif et probabiliste.....   | 112 |
| <b>Figure 60:</b> Formulaire de connexion au système. Une fois l'utilisateur authentifié toutes les interrogations et décisions de ce dernier seront stockées avec son identifiant.....  | 112 |
| <b>Figure 61:</b> Formulaire de choix de l'ontologie ou de la base de connaissances à utiliser.....  | 113 |
| <b>Figure 62 :</b> Les requêtes de sélection des catégories de signes. Ces catégories sont les superclasses des classes (signes) de l'ontologie. Elles peuvent contenir d'autres catégories, par exemple la catégorie « Lesion » contient les catégories « Bone_Lesion », « Lytic_lesion », etc..... | 113 |
| <b>Figure 63:</b> Formulaire adaptatif ou d'interrogation, ce formulaire est fonction de chaque base de connaissances ou ontologie.....  | 113 |
| <b>Figure 64:</b> Formulaire des résultats.....  | 114 |
| <b>Figure 65:</b> Prévalences calculées à partir de la base de cas des urgences odontologiques. ..   | 115 |
| <b>Figure 66:</b> Extrait des effectifs pour le calcul des sensibilités et des spécificités des signes pour les maladies.....  | 116 |
| <b>Figure 67:</b> Extrait des sensibilités et des spécificités calculées pour différentes maladies. ..   | 116 |
| <b>Figure 68:</b> Illustration d'un triplet sous forme d'un réseau de relations signe-diagnostic (hypothèse d'une maladie).....  | 122 |
| <b>Figure 69:</b> Tableau des effectifs utilisés pour le calcul des sensibilités et des spécificités des signes pour les maladies.....   | 125 |

# Liste des tableaux

|  |     |
|--|-----|
| <b>Tableau 1:</b> Tableau de contingence .....   | 31  |
| <b>Tableau 2:</b> Les 4 niveaux d'opérationnalisation des maladies. ....   | 56  |
| <b>Tableau 3:</b> Quelques relations de l'ontologie du NCIT .....  | 68  |
| <b>Tableau 4:</b> Prévalences de 4 types de lésions avec les sensibilités et spécificités (se/sp) des signes cliniques pour les types de lésions. .... | 72  |
| <b>Tableau 5:</b> Liste des constructeurs de la logique de description utilisée. ....  | 73  |
| <b>Tableau 6 :</b> Abréviations pour les schémas .....   | 76  |
| <b>Tableau 7:</b> Tableau des nouvelles relations de l'ontologie opérationnalisée. ....  | 79  |
| <b>Tableau 8:</b> Tableau des nouvelles relations utiles pour la représentation des probabilités dans l'ontologie. ....                                | 81  |
| <b>Tableau 9:</b> Tableau des nouvelles relations utiles pour les règles ajoutées dans l'ontologie opérationnalisée. ....                              | 83  |
| <b>Tableau 10:</b> Le dictionnaire de données. ....  | 96  |
| <b>Tableau 11:</b> Vue partielle de l'ensemble des triplets (r,f,d). ....  | 111 |
| <b>Tableau 12:</b> Vue partielle de l'ensemble des quintuplets (r,f,d,tv,u). ....  | 111 |
| <b>Tableau 13:</b> Vue partielle de l'ensemble des septuplets (f,d,c,prev,Se,Sp,LrMinus). ....   | 111 |
| <b>Tableau 14:</b> Vue partielle de l'ensemble des septuplets (f,d,c,prev,Se,Sp,LrPlus). ....  | 111 |



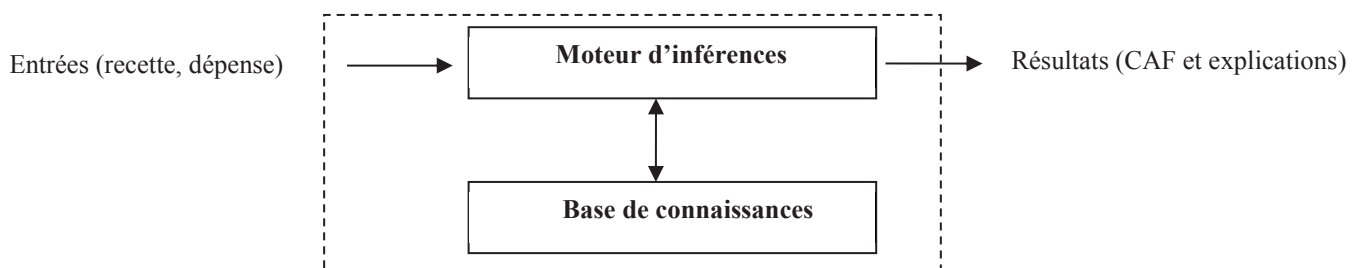
# Chapitre 1

## Introduction générale

### 1.1 Contextes et enjeux

Les Systèmes d'Aide à la Décision (SAD) (Figure 1) sont de manière générale des logiciels qui fournissent des informations qui permettent de prendre des décisions. Selon Steven Alter (Alter, 1980) : « *Ce sont des logiciels flexibles et interactifs qui aident les décideurs dans les processus d'extraction d'informations utiles pour identifier des problèmes et les résoudre, et pour prendre des décisions* ». Les SAD sont composés en général d'une base de connaissances et d'un moteur de raisonnement, dit aussi moteur d'inférences. La base de connaissances contient toutes les connaissances qui sont utiles pour la résolution des problèmes. Le moteur d'inférences quant à lui est un programme qui à partir d'un cas décrit (par exemple, le chiffre d'affaire du mois et les dépenses prévisionnelles du mois), exploite les connaissances contenues dans la base de connaissances pour extraire les résultats relatifs au cas (par exemple, la capacité d'autofinancement (CAF)). Les SAD nécessitent en particulier une représentation adéquate des connaissances mises en jeu, ainsi que des mécanismes efficaces d'exploitation de ces connaissances, ou de raisonnement.

Dans un SAD, le raisonnement est en général conditionné par le but que l'on souhaite atteindre pour résoudre un problème donné. Ainsi, il ne déduit pas l'ensemble des connaissances, mais uniquement celles qui sont associées au but recherché. Différents modes de raisonnement tels que déductif, abductif, inductif et par analogie sont utilisés pour produire des résultats (Arocha, Wang, & Patel, 2005).



**Figure 1: Structure de base d'un système d'aide à la décision. Des données sont fournies au système et le moteur d'inférences renvoie les résultats et les explications relatives à son raisonnement à partir du contenu de la base de connaissances.**

La médecine a été un champ d'application privilégié des systèmes d'aide à la décision. En effet, la simulation du diagnostic médical a été l'une des tâches réalisées par les premiers systèmes d'aide à la décision. Le diagnostic médical est une démarche par laquelle le médecin va déterminer l'affection dont souffre le patient, à partir des symptômes et des signes que ce dernier présente, et à l'aide d'éventuelles investigations complémentaires (Arocha et al., 2005; Bertaud-Gounot, 2008). Cette démarche est une tâche complexe qui repose sur des connaissances du domaine. Les connaissances du domaine sont les connaissances relatives (par exemple les relations entre les maladies et les signes) au domaine d'application (par exemple la cancérologie) et nécessaires pour que les méthodes de raisonnement puissent

s'exécuter (Charlet, 2003). Ces connaissances du domaine permettent d'une part de mettre en correspondance les observations issues de l'examen médical avec les critères associés aux différentes pathologies, et d'autre part de raisonner sur la présence ou l'absence de ces critères afin de restreindre la liste des diagnostics possibles et d'étendre celle des diagnostics exclus. De nombreuses formes de modélisation ont été proposées (par exemple, les modèles probabilistes, les règles, les réseaux sémantiques, les *frames*, etc.) afin d'être exploitées par des systèmes informatiques à des fins d'aide à la décision médicale et/ou thérapeutique, etc.

En effet, les systèmes d'aide au diagnostic médical grâce à la puissance de calcul et aux capacités mémoires des ordinateurs facilitent le traitement des grandes masses de connaissances médicales. Ces systèmes permettraient ainsi l'amélioration de la qualité des soins et de la prise en charge des malades par la réduction des erreurs médicales et par la détection des omissions (Seroussi, Le Beux, & Venot, 2013; Zhang, Patel, Turley, & Johnson, 2005).

Des années 60 à la fin des années 90 plusieurs tentatives ont été faites, on retiendra des systèmes comme :

- ❖ le « *Leeds abdominal pain system* », un système d'aide au diagnostic des patients souffrant de douleurs abdominales aiguës (De Dombal, Leaper, Staniland, McCann, & Horrocks, 1972; J. Horrocks, McCann, Staniland, Leaper, & De Dombal, 1972) s'appuyant sur un modèle probabiliste.
- ❖ « *ILIAD* », un système à base de réseau bayésien pour l'aide au diagnostic en médecine interne (Murphy et al., 1996; Warner et al., 1988).
- ❖ l'encyclopédie médicale « *ADM* » (Aide au Diagnostic Médical), un système à base de modèle relationnel permettant de faire de l'aide au diagnostic sur plus de 15000 pathologies toutes spécialités confondues (Seka, Pouliquen, & Le Beux, 1996).
- ❖ « *MYCIN* », un système à base de règles pour les diagnostics et les recommandations thérapeutiques des maladies infectieuses (Buchanan & Shortliffe, 1984; Van Melle, 1984).
- ❖ « *CASNET* », un système à base de réseau sémantique pour les diagnostics des glaucomes (Kulikowski & Weiss, 1982; Weiss, Kulikowski, & Safir, 1977).
- ❖ « *CENTAUR* », un système à base de *frames* pour l'interprétation des diagnostics des patients souffrant de pathologies respiratoires (Aikins, 1980).

Ces systèmes ont été plus ou moins bien accueillis dans les débuts, mais des études ont révélé quatre grandes catégories d'insuffisances (Bruaux, 2007; Clancey, 1993; Grigorova & Nikolov, 2008; Negnevitsky, 2002a; Page, 1990; Seroussi et al., 2013; Winograd, 1975) :

- ❖ La présentation des résultats sous forme de diagnostics probables sans explications n'a pas convaincu les médecins pour le cas des systèmes probabilistes, ainsi que le fait que le système ne soit pas capable de dire de manière explicite « pourquoi<sup>1</sup> » et « comment<sup>2</sup> » il a produit tels ou tels résultats pour les systèmes à base de règles (Clancey, 1993). Dans les systèmes à base de *frames* certaines opérations (comme des additions, multiplications, etc.) ne peuvent être faites que via des programmes informatiques externes écrits en LISP ou en C. Ce type de connaissances est alors

---

<sup>1</sup> En donnant la liste des signes observés qu'ils soient présents ou absents.

<sup>2</sup> En fournissant la nature des relations entre les signes qui impliquent une hypothèse de maladie donnée et cette maladie.

encapsulées dans des programmes informatiques et rendent ainsi les justifications des inférences incomplètes (Grigorova & Nikolov, 2008; Winograd, 1975).

- ❖ La sémantique du domaine est manquante dans les systèmes probabilistes, les systèmes à base de règles et les systèmes à base de réseaux bayésiens. Dans les systèmes probabilistes et réseaux bayésiens, ne sont représentées que des relations signes et maladies probabilisées, sans aucune autre explication. Les systèmes à base de règles ne modélisent que les connaissances pratiques de l'expert du domaine sans représentation des connaissances théoriques du domaine d'application (Page, 1990). Les modèles à base de *frames* quant à eux n'intègrent pas la distinction entre les conditions nécessaires et les conditions nécessaires et suffisantes (Negnevitsky, 2002a), ce qui entraînent des difficultés de maintenance de la cohérence du modèle pour des domaines complexes.
- ❖ Les modèles de connaissances élaborés via ces approches ne sont pas réutilisables pour certaines (probabiliste, règles et modèle relationnel) et difficilement réutilisables pour d'autres (réseaux sémantiques et *frames*). Les modèles probabilistes et les réseaux bayésiens utilisent des probabilités conditionnelles et les probabilités *a priori* qui sont toutes rattachées au contexte d'utilisation (Seroussi et al., 2013). Les modèles à base de règles ne tiennent compte que des connaissances pratiques de l'expert et n'ont pas de connaissances théoriques du domaine (Bruaux, 2007; Page, 1990). Les modèles à base de réseaux sémantiques n'intègrent pas la standardisation des relations (Grundstein, De Bonnières, & Para, 1988a).
- ❖ L'intégration au système d'information hospitalier (SIH) n'était pas effective et cela entraînait des difficultés d'utilisation (saisie importante des signes et symptômes du patient). Impossibilité d'accéder aux antécédents du patient via ces systèmes.

Ces dernières années avec l'avènement du web sémantique, l'utilisation des ontologies comme formalisme de représentation des connaissances a eu une progression significative dans de nombreux domaines, dont la médecine. Selon Thomas Gruber : « *Une ontologie est une spécification formelle et explicite d'une conceptualisation* » (Gruber, 1993). Cela signifie qu'une ontologie est une représentation des concepts et des liens entre les concepts d'un domaine. Un concept est une représentation générale et abstraite de la réalité d'un objet, d'une situation ou d'un phénomène. Les ontologies présentent de nombreux avantages, comme prendre en compte la sémantique du domaine étudié, faciliter la maintenance de la base de connaissances, faciliter la réutilisation des modèles existants. Elles permettent aussi d'assurer l'interopérabilité sémantique entre systèmes, d'organiser les entrepôts de données du domaine biomédical, de faciliter l'intégration des données entre systèmes, de faciliter l'aide à la décision et d'accéder aux informations pertinentes sur le Web (Bodenreider, 2008).

Le langage de représentation des ontologies recommandé par le « *World Wide Web Consortium* » (W3C) et le plus utilisé est le « *Web Ontology Language* » (OWL) (Cardoso, 2007; McGuinness & Harmelen, 2004). Certaines limites de OWL sont comblées par l'utilisation des langages de règles (Golbreich, Dameron, Gibaud, & Burgun, 2003) comme le « *Semantic Web Rule Language* » (SWRL) (I. Horrocks et al., 2004).

Des tentatives de mise en place de systèmes d'aide à la décision à base d'ontologies ont été faites, par exemple, le système « *Ontology Driven Differential diagnosis based on logical inference and probabilistic refinements* » (ODDIN) (García-crespo, Rodríguez-González, Mencke, Gómez-berbís, & Colomo-palacios, 2010) pour les diagnostics des patients en

médecine interne. ODDIN est encore au stade expérimental et les évaluations (García-crespo et al., 2010) présentent des résultats intéressants (cf. 3.3.4). D'autres travaux concernant les ontologies ont été menés, tels que :

- ❖ la représentation des connaissances imprécises et vagues (Fernando Bobillo, 2008; Peelen, Klein, Schlobach, Keizer, & Peek, 2007).
- ❖ la représentation des connaissances probabilistes (Bertaud-Gounot, 2008; Fenz, 2012).
- ❖ la représentation des pathologies multi-niveaux (Rodríguez-González, Torres-Niño, Mayer, Alor-Hernandez, & Wilkinson, 2012; Rodríguez-González et al., 2011; Rodríguez-González, Labra-Gayo, et al., 2012).
- ❖ la représentation des signes négatifs (Bertaud-Gounot, 2008; Ceusters, Elkin, & Smith, 2006, 2007).
- ❖ la prise en charge du raisonnement abductif (cf. 2.3.3) (Henson, Thirunarayan, Sheth, & Hitzler, 2011).

Ces systèmes et ces travaux montrent que l'utilisation des ontologies pour les systèmes d'aide au diagnostic médical est une voie prometteuse.

## 1.2 Objectifs de la recherche

Cette thèse a pour objectif principal de proposer des méthodes générales de modélisation des signes dans les ontologies biomédicales et la mise en place des mécanismes de raisonnement associés pour l'aide au diagnostic médical. Elle se déroule au sein de l'unité médicale de recherches « UMR936 » de l'université de Rennes 1 et s'inscrit à la suite d'autres travaux de cette unité de recherches, notamment les travaux de Charles Le Moing (Le Moing, 2009) et de Bertaud-Gounot (Bertaud-Gounot, Richard, Le Moing, & Duvauferrier, 2010; Bertaud-Gounot, 2008). Ces précédents travaux ont montré l'intérêt de l'utilisation des ontologies biomédicales pour les systèmes d'aide au diagnostic. Ils ont aussi mis en évidence : (a) la nécessité de la prise en compte du raisonnement « abductif » (cf. 2.3.3), (b) l'intérêt de la prise en compte des probabilités dans les ontologies, (c) la nécessité de la mise en place d'un moteur d'inférences gérant plusieurs modes de raisonnement (par exemple, le mode probabiliste, abductif et déductif, etc.) et capable de proposer de nouvelles observations indispensables pour affiner la liste des diagnostics possibles dans lesdits systèmes. Sur la base de cela, nous allons décrire :

- ❖ des travaux concernant les ontologies présents dans la littérature (cf. 3.3.1).
- ❖ les possibilités et les limites des langages standards de représentation du web sémantique, notamment OWL (cf. 3.3.2.2) et SWRL (cf. 3.3.2.3).
- ❖ des méthodes générales pour enrichir une ontologie biomédicale permettant de :
  - représenter les connaissances quantitatives (par exemple, une température inférieure à 37,6° C, une quantité d'immunoglobuline inférieure à 3g/dL) (cf. 4.2.1.2).
  - représenter les connaissances imprécises ou mal définies (par exemple, une fièvre forte, une hypercalcémie, etc.) (cf. 4.2.1.1).
  - représenter les signes négatifs, par exemple (pas de douleur lombaire, absence de fièvre, absence d'hypercalcémie, etc.) (cf. 4.2.2).
  - représenter la probabilité de la survenue d'un signe pour une maladie (cf. 4.2.1.3).

- représenter la probabilité *a priori* d'une maladie pour une population donnée ou selon un contexte. Le contexte serait par exemple, les patients âgés de plus de 50 ans ou les patients vivant dans la région de la Bretagne, etc. (cf. 4.2.1.3).

Et pour finir, nous allons proposer un prototype de système d'aide au diagnostic médical pour les tests et les évaluations des méthodes mises en place.

### 1.3 Plan du mémoire

Le contenu de ce mémoire est organisé en 6 chapitres. Le chapitre 2 fournit des définitions : (a) des notions du signe et du symptôme, (b) des notions de la maladie et du syndrome, (c) du diagnostic médical et (d) de la sémiologie quantitative.

Le chapitre 3 établit un état de l'art sur la représentation des connaissances et les ontologies d'une part, et d'autre part, il présente les objectifs détaillés de notre projet. Nous présentons la problématique de cette thèse et notre proposition. Nous insistons sur les méthodes de modélisation des connaissances élaborées et le système d'aide à la décision à base d'ontologie « *Ontology driven expert system* » (ODES) mis en place pour les tests et les évaluations desdites méthodes.

Le chapitre 4 présente le matériel et l'approche utilisés pour l'élaboration des méthodes de modélisation des connaissances et la mise en place du système ODES.

Le chapitre 5 présente les résultats, et il est divisé en 4 parties. Dans la première partie, nous présentons les étapes de construction d'une ontologie sémiologique biomédicale. Dans la deuxième partie, nous présentons les modes d'inférences possibles à partir de ces méthodes de modélisation. Dans la troisième partie, le système est présenté, et dans la dernière partie, nous montrons les résultats de l'évaluation desdites méthodes.

Le chapitre 6 positionne notre travail par rapport aux travaux existants en recensant ses apports et ses limites d'une part, et d'autre part, il présente la conclusion de ce travail et les perspectives.

Ce mémoire comporte également trois annexes : L'**Annexe 1** présente une classe « Java » de chargement de l'ontologie OWL pour le processus d'instanciation automatique. L'**Annexe 2**, présente une classe « Java » de chargement des ensembles de triplets dans la base de données relationnelle. Enfin, dans l'**Annexe 3**, nous présentons une classe « Java » de chargement des restrictions de l'ontologie en mémoire pour le moteur d'inférences du système.





## Chapitre 2

### Définitions

#### 2.1 Le signe et le symptôme

En médecine humaine, un « signe » ou un « ensemble de signes » peut être la manifestation objective d'une pathologie, relevée par le médecin (Bertaud-Gounot, 2008). Un signe peut apparaître à des niveaux d'observation différents (par exemple clinique, para-clinique ou complémentaire), impliquant des techniques de recueil plus ou moins rapides et/ou coûteuses (Fagot-Largeault, 2010). Les signes peuvent être scindés comme suit :

- ❖ Les signes cliniques sont des signes recueillis sans instrument lourd, par exemple le stéthoscope, le thermomètre, le tensiomètre, etc. Les signes cliniques peuvent être scindés en :
  - Signes généraux (par exemple, la température, la fatigue, etc.).
  - Signes focaux (délimité à une zone).
  - Signes fonctionnels ou symptômes. Ce sont des signes récupérés par l'interrogatoire, non vérifiable par un autre signe clinique, on peut citer la douleur, ou la dyspnée<sup>3</sup>.
  - Signes physiques. Ce sont ceux qui peuvent se vérifier à l'examen clinique. Par exemple, une rougeur, un bruit anormal au stéthoscope, etc.
- ❖ Les signes para-cliniques. Ceux qui sont observables via la radiographie, le scanner, etc.

Un symptôme représente une des manifestations subjectives d'une maladie ou d'un processus pathologique, tel qu'exprimé par le patient (Bertaud-Gounot, 2008). Les symptômes peuvent être multiples pour une pathologie donnée.

Une autre catégorisation des signes et symptômes existe et s'appuie sur la relation entre ces derniers et le diagnostic que veut établir un médecin. Cette catégorisation est fondée sur le postulat que chaque signe pris séparément n'a pas de sens en soi, et c'est plutôt sur un ensemble de signes que sera établi le diagnostic. Cette catégorisation est la suivante : signes ou symptômes positifs et signes ou symptômes négatifs. Dans les deux sous-sections (2.1.1 et 2.1.2) suivantes nous utilisons le terme « signe » pour désigner indifféremment le signe en lui-même ou le symptôme.

##### 2.1.1 Le signe positif

Un signe est dit positif si c'est sa présence qui a une signification, par exemple une quantité d'immunoglobuline de type A supérieure à 3g/dL est un signe positif pour le myélome multiple (Rajkumar, 2011).

---

<sup>3</sup> C'est une difficulté respiratoire.

### 2.1.2 Le signe négatif

Werner Ceusters (Ceusters et al., 2006, 2007) définit les « signes négatifs » ou « signes absents » comme étant tous les signes qui sont exprimés avec des négations, par exemple, « pas de douleur lombaire, absence de fièvre etc. ». C'est aussi, dans certaines situations un signe dont c'est l'absence qui a une signification, par exemple, l'absence de CRAB (C- hypercalcémie, R- insuffisance rénale, A- anémie, B- lésion osseuse) est caractéristique de la gammapathie monoclonale à signification indéterminée (MGUS) (Rajkumar, 2011).

## 2.2 La maladie et le syndrome

Nous présentons dans cette partie trois définitions de la maladie, celle de l' « *Ontology for General Medical Science* » (OGMS) (Scheuermann, Ceusters, & Smith, 2009), celle de Émile Littré et celle de Anne Fagot-Largeault (Fagot-Largeault, 2010).

En effet, selon l'OGMS (Scheuermann et al., 2009) : « *Une maladie est une disposition à subir un processus pathologique causé par un ou plusieurs désordres sous-jacents* ». En d'autres termes, la maladie est une disposition réalisée dans des processus pathologiques. Nous utilisons le terme « réalisation » pour faire référence au processus par lequel une disposition est réalisée. Chaque disposition dans l'organisme a un site physique. Le site physique d'une maladie est une certaine combinaison de composants physiques dans l'organisme. Quel que soit l'organisme, il y a une large variété de dispositions, certaines sont associées à la bonne santé, d'autres à la maladie. Par exemple, le SIDA est une disposition du patient sain à subir une persistance et une duplication pathogène (processus pathologique) suivis des infections opportunistes survenant à cause de la faiblesse du système immunitaire (site physique). Une disposition est une fonction d'un organisme en vertu de laquelle elle peut initier des processus spécifiques quand certaines conditions sont satisfaites. Par exemple, la disposition du corps humain à avoir besoin d'eau en cas de déshydratation, la disposition du corps humain à avoir des hausses de température en cas d'infection. Le processus pathologique est un processus physique qui est une manifestation d'un désordre. Un désordre est une combinaison causale des composants physiques qui sont cliniquement anormaux (Scheuermann et al., 2009).

Selon Émile Littré (Bertaud-Gounot, 2008) : « *La maladie est une altération organique ou fonctionnelle de la santé considérée dans son ensemble comme une entité définissable* ».

Et enfin, selon Anne Fagot-Largeault (Fagot-Largeault, 2010) : « *La maladie est réduite à l'ensemble des symptômes et des signes par lesquels elle se manifeste, son nom est à ce moment-là une étiquette posée sur une constellation (éventuellement disparate) de symptômes et de signes* ». Par exemple, une gammapathie monoclonale à signification indéterminée (MGUS) est caractérisée par une quantité d'immunoglobuline de type A inférieure à 3 g/dL et une absence de CRAB (cf. 2.1.2).

Le syndrome est une collection de signes ou de symptômes survenant simultanément sans étiologie identifiée (Bertaud-Gounot, 2008).

Pour la suite, la définition de la maladie que nous retenons est celle de Anne Fagot-Largeault (Fagot-Largeault, 2010).

## 2.3 Le diagnostic médical

Nous avons dit plus haut (cf. 1.1) que le diagnostic médical est une démarche par laquelle le médecin va déterminer l'affection dont souffre le patient, à partir des symptômes et des signes que ce dernier présente, et à l'aide d'éventuelles investigations complémentaires (Arocha et al., 2005; Bertaud-Gounot, 2008). Cette démarche repose sur la recherche des causes (étiologie) et des effets (signes et symptômes) de l'affection (Bertaud-Gounot, 2008). Randolph Miller (Miller, 2009) définit le diagnostic médical comme : « *l'acte d'associer le nom d'une ou plusieurs maladies ou syndromes à des manifestations observées (antécédents, symptômes, signes) sur un patient* ».

C'est un processus qui résulte de la confrontation d'un problème réel à l'expérience acquise et à un corpus de connaissances théoriques. Il se déroule en trois étapes : (a) l'enquête interrogative, (b) l'examen clinique proprement dit et (c) les examens complémentaires (Masquelet, 2006a).

- ❖ Premièrement, le médecin prend connaissance des symptômes se manifestant chez un patient. A partir des symptômes, il formule des hypothèses diagnostiques initiales.
- ❖ Dans un deuxième temps, il procède à un examen initial du patient qui lui permet d'augmenter la part de confiance en certaines hypothèses, et de la diminuer pour d'autres.
- ❖ Le médecin obtient des signes supplémentaires via des examens biologiques ou d'imageries. En même temps, il pose au patient des questions dont les réponses peuvent être utiles à conforter ou rejeter une hypothèse initialement formulée.

Le médecin « réalise » une mise en correspondance entre les informations obtenues au cours des trois étapes précédentes avec les connaissances qu'il possède de par sa formation et son expérience. Si au terme des étapes précédentes, le taux de confiance d'une certaine hypothèse s'accroît au point de dissiper le doute sur la maladie à laquelle est confronté le médecin, ce dernier peut alors formuler son diagnostic final et prescrire le traitement adéquat au patient. Si le cas reste ambigu après les trois étapes indiquées, le médecin cherche alors une autre source d'informations qui puisse apporter une quantité d'informations supplémentaires permettant d'éliminer l'ambiguïté. Souvent, il demande un examen complémentaire qui peut être sous forme d'analyses biologiques, d'imagerie médicale, etc. Il acquiert des informations supplémentaires qui viennent compléter la quantité d'informations dont il dispose déjà, et qui lui permet de confirmer ou d'infirmer la ou les hypothèses qu'il a déjà faites. Si le médecin n'arrive toujours pas à établir un diagnostic, une dernière étape consiste à ce qu'il ait recours à l'étude d'une base de cas similaires traités par le passé (son expérience) afin d'établir une correspondance avec le cas actuel auquel il est confronté en s'appuyant sur toutes les informations dont il dispose. Il utilise alors les cas les plus similaires (leurs solutions) afin d'en extraire des informations susceptibles de l'aider à trouver une solution à son cas.

Le processus du diagnostic médical peut relever d'un raisonnement déductif<sup>4</sup>, inductif, abductif, par analogie, etc. (Patel, Arocha, & ZHANG, 2004; Pottier & Planchon, 2011; Seroussi et al., 2013).

### 2.3.1 Le raisonnement déductif

Le raisonnement déductif<sup>5</sup> va du général au particulier ou du principe à la conséquence et permet de déduire des conclusions valides à partir d'un ensemble de prémisses, c'est une approche descendante (Degoulet & Fieschi, 1994a; Masquelet, 2006b). Par exemple,

- ❖ Toutes les rhinopharyngites virales ont une évolution spontanément favorable (règle),
- ❖ Ce patient a une rhinopharyngite virale (cas).
  - Ce patient aura une évolution spontanément favorable (résultat).

Ce mode de raisonnement peut être appliqué via un chaînage avant, un chaînage arrière ou un chaînage mixte suivant les règles du « *modus ponens* » et du « *modus tollens* ».

#### Modus ponens

|                                   |    |                             |
|-----------------------------------|----|-----------------------------|
| Si A est vrai                     |    | A                           |
| Et A implique C, alors C est vrai | ou | $\frac{A \rightarrow C}{C}$ |

#### Modus tollens

|  |    |                                  |
|--|----|----------------------------------|
| Si $\neg C$ est vrai                     |    | $\neg C$                         |
| Et A implique C, alors $\neg A$ est vrai | ou | $\frac{A \rightarrow C}{\neg A}$ |

Le « chaînage avant » est un mode de raisonnement qui part des faits et règles de la base de connaissances, et tentent de s'approcher des conclusions recherchées par le problème.

Le « chaînage arrière » quant à lui part des hypothèses recherchées par le problème, et tente par l'intermédiaire des règles, de « remonter » à des faits connus.

Le « chaînage mixte » est une combinaison du chaînage avant et du chaînage arrière.

### 2.3.2 Le raisonnement inductif

Le raisonnement inductif montre une proposition générale à partir d'observations particulières. Il est utilisé dans les processus d'apprentissage et est ascendant, c'est-à-dire qu'à partir d'exemples, on produit des règles qui décrivent ces exemples. Si ces règles sont bonnes, elles s'appliquent non seulement à ces exemples particuliers mais aussi d'une manière générale à d'autres cas (Degoulet & Fieschi, 1994a; Masquelet, 2006b). Par exemple,

- ❖ Un groupe de patients a une rhinopharyngite virale (cas),
- ❖ Tous les patients ont guéri spontanément (résultat),
- ❖ Toutes les rhinopharyngites virales ont une évolution spontanément favorable (règle).

<sup>4</sup> Mais pas toujours et on verra dans les sous-sections suivantes qu'il est abductif (cf. 2.3.3).

<sup>5</sup> C'est le mode de raisonnement mis en œuvre par les moteurs d'inférences du web sémantique (cf. 3.3.2.7).

### 2.3.3 Le raisonnement abductif

Le raisonnement abductif<sup>6</sup> est une forme d'inférence qui part des données (prémisses) décrivant quelque chose vers les hypothèses (conclusion) permettant d'expliquer ces données. Par exemple, étant donné une proposition  $a$ , on essaie de savoir ce qui peut l'avoir induite. Pour cela, on regarde s'il y a une implication de la forme  $b \rightarrow a$  et, si c'est le cas, on établit que  $b$  est la « cause » de  $a$ .

Dans le cadre de la médecine, selon Douglas Walton (Walton, 2004), Charniak et McDermott (Josephson, Josephson, & Tanner, 1996) le diagnostic médical est une forme de raisonnement abductif car il permet d'inférer les causes d'un phénomène, d'un dysfonctionnement, d'un signe, d'un symptôme etc. Il s'appuie sur le modèle suivant (Josephson et al., 1996) :

Soit D une collection de données (faits, observations, signes et symptômes),  
Soit H une hypothèse de D (si H est vrai, il explique D) et  
Aucune autre hypothèse ne peut mieux expliquer D que H  
Alors, l'hypothèse H est probablement vrai

Le raisonnement abductif génère des hypothèses permettant d'expliquer la présence des signes ou des symptômes et il donne la possibilité de générer des hypothèses même si on n'a pas observé tous (de manière simultanée) les signes ou symptômes caractérisant ces hypothèses (Pottier & Planchon, 2011). Mais, il ne se limite pas seulement à la génération d'hypothèse, il permet aussi d'évaluer ces hypothèses. C'est un raisonnement incrémental qui pour chaque signe ou symptôme observé peut confirmer ou infirmer une hypothèse précédemment générée (Josephson et al., 1996).

### 2.3.4 Le raisonnement par analogie

Le raisonnement par analogie est le procédé qui consiste à affecter une ou plusieurs propriétés à un objet (la cible) sachant qu'un objet de référence (la base) l'a ou les possède déjà. Ce transfert peut nécessiter ou induire l'apparition d'autres propriétés sur la cible. L'analogie exploite des relations de correspondances et des relations de dépendances afin de transposer des connaissances d'un univers à un autre. L'analogie se fonde sur la notion de similitude de termes dans un certain univers, du type : A est à B, ce que C est à D. Le raisonnement à partir de cas en est un sous-type (Aamodt & Plaza, 1994).

## 2.4 La sémiologie quantitative

La sémiologie médicale est la science des signes. Elle s'intéresse : (a) aux manifestations d'une maladie que peut relever le médecin (signes physiques, fonctionnels, généraux) ou celles engendrant une plainte du patient (symptômes), (b) à la façon de les relever (interrogatoire, examen clinique), et (c) de les présenter (observation, tableau, syndromes, etc.) afin de poser un diagnostic (Bertaud-Gounot, 2008).

---

<sup>6</sup> Ce mode de raisonnement n'est pas pris en charge par les moteurs d'inférences du web sémantique car ces derniers sont tous déductifs (cf. 3.3.2.7).

La sémiologie quantitative est une partie de la sémiologie médicale, qui a pour rôle de quantifier la pertinence des signes pour des maladies données dans le but d'établir des diagnostics fiables (Chatellier, Ménard, & Degoulet, 1996). On utilise en particulier dans ce domaine les termes de prévalence (Prev), de sensibilité (Se), de spécificité (Sp), de rapport de vraisemblance positif (RVP) et de rapport de vraisemblance négatif (RVN), de valeur prédictive positive (VPP), de valeur prédictive négative (VPN) et de probabilité post-test (PPT).

**La prévalence** est la probabilité d'avoir une maladie dans une population donnée (par exemple, la population des hommes âgés de plus de 50 ans). La prévalence est aussi appelée probabilité *a priori*.

**La sensibilité** d'un signe pour une maladie est la probabilité qu'un patient ait un signe s'il a la maladie.

**La spécificité** d'un signe pour une maladie est la probabilité qu'un patient n'ait pas un signe s'il n'a pas la maladie.

**Le rapport de vraisemblance positif** est le rapport de la probabilité qu'un patient malade ait le signe (c'est-à-dire la sensibilité du signe) et de la probabilité que le signe soit présent chez les non-malades.

**Le rapport de vraisemblance négatif** est le rapport de la probabilité que le signe soit absent chez les malades et de la probabilité que le signe soit absent chez les non-malades (c'est-à-dire la spécificité).

**La valeur prédictive positive** est la probabilité d'avoir une maladie donnée si le patient a le signe (Degoulet & Fieschi, 1994b).

**La valeur prédictive négative** est la probabilité que la maladie soit absente si le patient n'a pas le signe (Degoulet & Fieschi, 1994b).

**La probabilité post-test** est la probabilité que la maladie soit présente ou absente si le patient a le signe ou si le patient n'a pas le signe, ou encore si le patient a des signes présents et des signes absents. Dans la suite les termes probabilités post-test, VPP et VPN seront utilisés de manière interchangeable.

### 2.4.1 Tableau de contingence

En pratique clinique, quatre situations peuvent être observées qui résultent de la combinaison de la présence ou non du signe et de la présence ou de l'absence de maladie. On a coutume de représenter ces situations sous formes d'un tableau (Masquelet, 2006b) (Tableau 1). On appelle VP les vrais positifs (le signe et la maladie sont présents), FP les faux positifs (le signe est présent, mais la maladie est absente), FN les faux négatifs (le signe est absent, mais la maladie est présente), VN les vrais négatifs (le signe et la maladie sont tous les deux absents).

Tableau 1: Tableau de contingence

|               | Malade | Non malade | Totaux       |
|---------------|--------|------------|--------------|
| Signe présent | VP (a) | FP (b)     | VP+FP        |
| Signe absent  | FN (c) | VN (d)     | FN+VN        |
| Totaux        | VP+FN  | FP+VN      | VP+FP+ FN+VN |

**Se** =  $\frac{VP}{(VP+FN)}$  : La sensibilité (Sn), ou constance du signe, est le pourcentage des sujets qui ont le signe (a) parmi ceux qui ont la maladie (a + c), et

**Sp** =  $\frac{VN}{(FP+VN)}$  : La spécificité (Sp), ou exclusivité du signe, est le pourcentage des sujets qui n'ont pas le signe (d) parmi ceux qui n'ont pas la maladie (b + d).

**Prev** =  $\frac{VP+FN}{(VP+FP+FN+VN)}$  : Le pourcentage de malade parmi sur la population totale étudiée.

$$\mathbf{VPP} = \frac{VP}{(VP+FP)} \quad \text{et} \quad \mathbf{VPN} = \frac{VN}{(FN+VN)}$$

Le calcul des VPP et VPN peut se faire aussi par application du théorème de Bayes (Aronoff, 2010; McGee, 2002). Cette application passe par l'utilisation des valeurs de la sensibilité, spécificité et prévalence.

$$\mathbf{VPP} = \frac{\mathbf{Se} \times \mathbf{Prev}}{((\mathbf{Se} \times \mathbf{Prev}) + ((1 - \mathbf{Sp})(1 - \mathbf{Prev})))} \quad (\text{Eq. 1})$$

$$\mathbf{VPN} = \frac{\mathbf{Sp} \times (1 - \mathbf{Prev})}{((\mathbf{Sp} \times (1 - \mathbf{Prev})) + (1 - \mathbf{Se})(\mathbf{Prev}))} \quad (\text{Eq. 2})$$

Cette méthode présente l'intérêt d'être simple de mise en œuvre. Elle est appropriée pour obtenir les VPN et VPP dans le cadre d'un signe et d'une maladie. Lorsque l'on recherche les VPP et VPN non plus d'un seul signe, mais d'un ensemble de signes présents et de signes absents, une autre méthode est plus commode. Elle fait intervenir les rapports de vraisemblance positifs (RVP) et négatifs (RVN) de chaque signe.

$$\mathbf{RVP} = \mathbf{SE} / (1 - \mathbf{SP}) \quad (\text{Eq. 3})$$

$$\mathbf{RVN} = (1 - \mathbf{SE}) / \mathbf{SP} \quad (\text{Eq. 4})$$

On appelle RV le rapport de vraisemblance positif ou négatif, selon que le signe observé est présent ou absent. Soit i1, i2, i3, ..., in les signes observés et n le nombre total de signes observés.

$$\mathbf{VPP} = \left( \frac{\mathbf{Prev}}{1 - \mathbf{Prev}} \right) \prod_{i=1}^n \mathbf{RVP}_i / (1 + \left( \frac{\mathbf{Prev}}{1 - \mathbf{Prev}} \right) \prod_{i=1}^n \mathbf{RVP}_i) \quad (\text{Eq. 5})$$

$$\mathbf{VPN} = \left( \frac{\mathbf{Prev}}{1 - \mathbf{Prev}} \right) \prod_{i=1}^n \mathbf{RVN}_i / (1 + \left( \frac{\mathbf{Prev}}{1 - \mathbf{Prev}} \right) \prod_{i=1}^n \mathbf{RVN}_i) \quad (\text{Eq. 6})$$



$$\text{PPT} = \left( \frac{\text{Prev}}{1-\text{Prev}} \right) \prod_{i=1}^n \text{RV}_i / \left( 1 + \left( \frac{\text{Prev}}{1-\text{Prev}} \right) \prod_{i=1}^n \text{RV}_i \right) \quad (\text{Eq. 7})$$

## 2.4.2 Interprétations

Un signe très sensible (Se proche de 100%) peut être absent chez des malades (faux négatifs, c), mais peut être présent chez certains non-malades (faux positifs, b). Dans ce cas, le signe est nécessaire mais non suffisant au diagnostic : tous les malades l'ont, mais certains non-malades aussi.

Un signe très spécifique (Sp proche de 100%) peut être présent chez des non-malades (faux positifs, b), mais peut être absent chez des malades (faux négatifs, c). Dans ce cas, le signe est suffisant mais non nécessaire au diagnostic : tous ceux qui n'ont pas la maladie n'ont pas le signe. On dit que le signe est pathognomonique (ou symptomatique) : le signe c'est la maladie, par exemple, le signe de Koeplich est pathognomonique pour la rougeole.

Les Se et Sp sont exportables car elles sont peu influencées par le contexte clinique où on les a établies pour un signe dans le diagnostic d'une maladie. Elles ne dépendent pas de la prévalence de la maladie.

Les VPP, VPN et PPT ont une signification concrète car elles indiquent la probabilité de la maladie chez un patient selon qu'il a ou n'a pas le signe. Cependant, elles dépendent du contexte clinique, donc de la prévalence, et sont par conséquent non-transposables.

## 2.5 Conclusion

Le diagnostic médical relève du raisonnement scientifique, mais il possède des difficultés spécifiques telles que l'incertitude et l'imprécision. Chaque étape de diagnostic médical peut donner lieu à des erreurs qui, dans le domaine de la santé peuvent être lourdes de conséquences. Pour éviter ces erreurs on peut se servir de la proposition de Zhang (Zhang, Patel, Johnson, & Shortliffe, 2004) qui consiste à mettre en place :

- ❖ Un guide de bonne pratique.
- ❖ Une représentation formelle des maladies. C'est-à-dire un modèle clair et précis des relations entre les maladies et les signes.
- ❖ Un système d'aide au diagnostic qui doit permettre d'avoir un affichage détaillé des informations nécessaires au processus de raisonnement (signe caractérisant une hypothèse de maladie, dossier médical du patient, antécédent, connexion au SIH, etc.).

Dans la suite de ce travail, nous nous focaliserons sur les travaux existants dans le domaine de la représentation des connaissances, les ontologies et les systèmes d'aide au diagnostic médical.

## Chapitre 3

### État de l’art

#### 3.1 La connaissance

Dans ce mémoire nous n’abordons le concept de « connaissance » que dans le cadre de l’informatique. La « connaissance » correspond à un ensemble d’informations organisées de manière à donner plus de sens (Zins, 2006). Par exemple, la règle « Si condition Alors conclusion » représente une connaissance élémentaire reliant deux informations de base : une information de condition et une information de conclusion (Zins, 2006).

#### 3.2 La typologie des connaissances

Les connaissances peuvent être classées par leur nature (De Jong, 1996; Grundstein, De Bonnières, & Para, 1988b), à savoir : déclaratives, procédurales et contrôles (Désilets, 1997; Grundstein et al., 1988a).

##### 3.2.1 Les connaissances déclaratives

Grundstein (Grundstein et al., 1988a) définit les « connaissances déclaratives » comme le « savoir théorique », le « quoi ». Elles sont définies sans ordre spécifique, sans dimension temporelle (par exemple, on peut dire que toutes les maladies plasmocytaires sont des gammopathies monoclonales). Elles sont constituées de concepts liés entre eux pour former des propositions (par exemple : la gammopathie monoclonale à signification indéterminée (MGUS) est un état précancéreux faisant partie des gammopathies monoclonales).

Celles-ci, cependant, ne permettent pas d’agir à elles seules sur le réel, en d’autres termes, l’utilisation de ces connaissances n’est pas spécifiée. Plutôt statiques, elles devront être traduites en conditions et procédures pour permettre une action.

##### 3.2.2 Les connaissances procédurales

Le savoir-faire constitue la deuxième catégorie des connaissances. Ce sont les procédures, la connaissance du « comment » de l’action (De Jong, 1996). Elles constituent la manière dont une connaissance est traitée, et elles traduisent à ce moment-là, le « comment ». Elles permettent de dérouler des séquences d’actions élémentaires (les éléments de ces séquences d’actions peuvent avoir une succession temporelle) et elles sont très liées au but qu’elles se proposent d’atteindre. C’est par elles qu’on apprend à réaliser une tâche réelle (par exemple, comment obtenir le diagnostic d’un patient à partir des signes observés ? comment dire qu’une instance donnée est membre d’une classe donnée ? analyser les interactions médicamenteuses dans une ordonnance).

Cependant, les connaissances déclaratives et les connaissances procédurales doivent avoir des conditions d’utilisation qui leur permettront d’être appelées au bon moment. Pour

cela, il faut les lier à une troisième catégorie de connaissances permettant de définir le « quand » et le « pourquoi » utiliser une connaissance. Cette troisième catégorie de connaissances constituent selon Clancey (Clancey, 1983) les connaissances de contrôle.

### 3.2.3 Les connaissances de contrôle

Clancey (Clancey, 1983) définit les « connaissances de contrôle » comme celles qui sont utilisées par le moteur d’inférences pour exploiter les connaissances du domaine (connaissances déclaratives et connaissances procédurales) afin de résoudre un problème. Bruno Bachimont (Bachimont, 1992) complète cette définition en ajoutant que : « *Les connaissances de contrôle sont des connaissances permettant d’utiliser des représentations de connaissances du domaine dans le système donné* ».

En effet, elles sont la médiation nécessaire entre une architecture (le système qui les utilise) et les connaissances (celles du domaine) qui doivent être prises en compte pour résoudre le problème.

Les connaissances de contrôle spécifient quand et comment il faut mener des traitements, par exemple, dériver les probabilités d’avoir une maladie donnée à partir des signes observés, faire des inférences, faire de l’acquisition de données, etc. Elles permettent dans différents contextes d’appliquer les connaissances déclaratives et procédurales (De Jong, 1996; McCormick, 1997). L’équivalent des connaissances de contrôle dans le système MYCIN (Buchanan & Shortliffe, 1984) sont les métarègles (Davis & Buchanan, 1984).

Enfin, pour que les connaissances soient utilisées par un système, elles doivent être représentées de manière claire, précise et sans ambiguïté dans un formalisme donné.

## 3.3 La représentation des connaissances

La représentation des connaissances est un domaine de recherche en intelligence artificielle qui a pour but de représenter et organiser formellement les connaissances pour les utiliser et les partager grâce à un ensemble de méthodes et d’outils. Dans une représentation formelle, les connaissances sont représentées par des objets logiques liés par des propriétés, axiomes et règles. Elle permet d’automatiser divers traitements sur ces connaissances.

Les ontologies constituent des outils formels permettant de représenter les connaissances des domaines complexes. Des systèmes d’aide au diagnostic médical à base d’ontologie ont été développés, nous nous intéressons au système informatisé d’aide au diagnostic en urgences odontologiques fait par Charles Le Moing (Le Moing, 2009) (cf. 3.3.3) et au système ODDIN réalisé par García-crespo (García-crespo et al., 2010) (cf. 3.3.4). Après avoir présenté ces systèmes, d’autres travaux sur les ontologies sont présentés.

Mais, avant cela, nous faisons une présentation générale des ontologies dans le domaine de l’intelligence artificielle et des différents langages et outils utilisés pour les mettre en œuvre et les manipuler.

### 3.3.1 Les ontologies

Selon Christopher Welty le terme « ontologie » a été emprunté à une branche de la philosophie que l’on appelle l’**Ontologie**, et qui est l’étude de l’être en tant qu’être ou encore l’étude de ce qui existe, c’est-à-dire l’ensemble des connaissances que l’on a sur le monde (Welty & Guarino, 2001).

Dans le domaine de l’intelligence artificielle, ce terme a une définition différente. Plusieurs définitions ont été proposées, nous allons en citer les plus pertinentes :

- ❖ Pour Robert Neches (Neches et al., 1991) : « *Une ontologie définit les termes et les relations de base comportant le vocabulaire d’un domaine aussi bien que les règles pour combiner des termes et les relations afin de définir des extensions du vocabulaire* ».
- ❖ Selon Borst Willem (Borst Willem, 1997) : « *Une ontologie est une spécification formelle d’une conceptualisation partagée* ».
- ❖ Rudi Studer (Studer, Benjamins, & Fensel, 1998) fait une combinaison des définitions de Thomas Gruber (cf. 1.1) (Gruber, 1993) et de Borst Willem (Borst Willem, 1997) : « *Une ontologie est une spécification formelle et explicite d’une conceptualisation partagée* ». Une conceptualisation réfère à un modèle abstrait d’un phénomène dans le monde. C’est une idée qu’on se fait de quelque chose. Explicite signifie que le type de concepts utilisés et les contraintes liées à leur usage sont définis explicitement. Formel stipule que l’ontologie doit être traduite en langage interprétable par une machine. Partagé signifie qu’une ontologie capture la connaissance consensuelle, c’est-à dire non réservée à quelques individus, mais partagée par un groupe ou une communauté d’acteurs.

Une ontologie est une liste finie de termes, et de relations entre ces termes. Ces termes correspondent à des concepts (classes d’objets) du domaine. Une ontologie est aussi un moyen de représenter des objets à des fins de manipulations informatiques.

Dans la suite nous utilisons la définition de l’« ontologie » de Rudi Studer (Studer et al., 1998).

#### 3.3.1.1 Applications des ontologies

Les ontologies jouent un rôle majeur dans les organisations et les systèmes en : (a) facilitant la communication inter-agent via le partage des connaissances (Dolia, 2010), (b) fournissant une base de connaissances pour les systèmes d’aide à la décision (Bodenreider, 2008), (c) en mettant à disposition un référentiel unique pour l’intégration des données entre des systèmes. Elles peuvent aussi être utilisées dans le domaine de la recherche documentaire et pour le traitement automatique du langage naturel (Cardoso, 2007; Haav & Lubi, 2001). De manière plus détaillée l’utilisation des ontologies offre de nombreux avantages (Bodenreider, 2008; Gruber, 1995; Rodriguez-Gonzalez et al., 2010) comme :

- ❖ fournir un cadre de représentation et de partage des connaissances via un vocabulaire commun. Elles fournissent un format d’échange des connaissances et un vocabulaire

spécifique pour la communication et elles facilitent l’interopérabilité (Bodenreider, 2008).

- ❖ faciliter la réutilisation des connaissances (Musen, 1992). Il est possible d’hériter des propriétés des concepts contenues dans des ontologies dites de haut niveau, pour la mise en place d’une ontologie de niveau inférieur. Par exemple, une ontologie de domaine et une ontologie de tâche ou d’application, etc.<sup>7</sup>.
- ❖ faciliter l’intégration des données (Bodenreider, 2008). Elles peuvent être utilisées comme référentiel pivot lors des transferts de données entre système, en ce sens qu’elles permettent de faire la correspondance entre les termes du système source et les termes de référence (ceux de l’ontologie) avant l’enregistrement dans le système destination.
- ❖ fournir une base de connaissances pour les systèmes de gestion de recrutement des patients éligibles pour des essais cliniques (Cuggia, Besana, & Glasspool, 2011; Stenzhorn et al., 2010). En effet, dans ce type de système, les ontologies sont des bases de connaissances contenant les critères d’éligibilités et les groupes correspondants à ces critères. Alors, une mise en correspondance est faite entre les données des patients et les critères d’éligibilités présents dans la base de connaissances. Le principe est le même pour les systèmes d’aide au diagnostic médical (Bertaud-Gounot, Duvauferrier, & Burgun, 2012; Bodenreider, 2008) car comme dans le cas du recrutement des patients pour les essais cliniques, les ontologies biomédicales sont utilisées comme des bases de connaissances. Les relations entre les maladies et les signes y sont représentées et le système fait une mise en correspondance entre les données ou signes du patient et le contenu de la base de connaissances.
- ❖ faciliter l’automatisation de l’acquisition des connaissances.
- ❖ faciliter le maintien de la séparation des connaissances déclaratives, procédurales et de contrôle.

### 3.3.1.2 Typologie des ontologies

Indépendamment des langages de représentation, on distingue différents niveaux d’ontologies selon le domaine modélisé et éventuellement les tâches pour lesquelles elles sont conçues (Burgun & Bodenreider, 2001).

- ❖ Les ontologies supérieures (top-level or upper-level ontologies : ULO) regroupent des catégories de haut niveau, c’est-à-dire des concepts généraux qui sont applicables dans tous les domaines tels que les « entités », les « événements », les « états », les « processus », les « actions », le « temps », « l’espace », les « relations », les « propriétés ». Les ontologies supérieures sont censées être universelles, c’est-à-dire qu’elles ne doivent pas faire référence à des domaines spécifiques. Elles doivent être multi-usages, c’est-à-dire qu’elles ne doivent pas avoir été conçues pour une tâche particulière.
- ❖ Les ontologies générales représentent des connaissances générales, indépendamment des domaines ou des tâches spécifiques, avec un niveau de précision moyen. Aussi

<sup>7</sup> Ces notions seront expliquées en détail dans les sections suivantes.

appelées, méta-ontologies ou core-ontologies, elles véhiculent des connaissances génériques moins abstraites que celles véhiculées par les ontologies supérieures, mais assez générales néanmoins pour être réutilisées à travers différents domaines. Elles peuvent adresser des connaissances factuelles (Generic domain ontology) ou encore des connaissances visant à résoudre des problèmes génériques (connaissances procédurales) appartenant à ou réutilisables à travers différents domaines (Generic task ontology). Deux exemples de ce type d’ontologie sont : (a) l’ontologie méréologique (Borst Willem, 1997) contenant des relations « partie-de » et (b) l’ontologie topologique contenant des relations « associé-à ».

- ❖ Les ontologies de domaine sont spécifiques d’un domaine (par exemple, le domaine médical, l’électronique, la mécanique, etc.) mais indépendantes d’une tâche particulière. Elles sont censées refléter la réalité sous-jacente et la théorie du domaine,
- ❖ Les ontologies d’application concernent un domaine limité et des tâches spécifiques. Elles font généralement partie d’une application et contiennent un nombre relativement faible de concepts qui devraient être définis de manière assez détaillée avec des relations et des règles d’inférences qui permettent de raisonner sur les concepts pour les tâches prévues. Cette ontologie est la plus spécifique. Les concepts dans l’ontologie d’application correspondent souvent aux rôles joués par les entités du domaine tout en exécutant une certaine activité (Maedche & Staab, 2001).
- ❖ Les ontologies de référence : Tandis que les ontologies d’application sont destinées à un usage spécifique, les ontologies de référence sont développées indépendamment de toute application de façon à refléter la réalité. Par définition, les ontologies de haut niveau sont des ontologies de référence car elles sont destinées à constituer la structure de haut niveau de nombreuses autres ontologies (de domaine ou d’application). Certaines ontologies de domaine peuvent aussi être des ontologies de référence (Burgun, 2006). Par exemple « *the Foundational Model of Anatomy* » (FMA) est une ontologie de référence de toute l’anatomie humaine. Elle a été proposée comme référence pour décrire la physiologie et la pathologie (Rosse & Mejino, 2003). Ces ontologies doivent permettre différents types d’applications comme la gestion de l’hétérogénéité des données par une normalisation sémantique des différents concepts, le raisonnement sur des entités complexes en rapport avec la réalité ou le raisonnement sur des entités individuelles en rapport avec des données.

D’autre part, les ontologies peuvent aussi être classées selon le niveau de complétude (Bachimont, 2000; Mizoguchi & Ikeda, 1998), c’est-à-dire le niveau de couverture de la sémantique du domaine étudié. Bruno Bachimont propose la classification sur trois niveaux (Bachimont, 2000) :

- ❖ Ontologie de niveau 1 - Sémantique : Tous les concepts (caractérisés par un terme/libellé) doivent respecter les quatre principes différentiels : 1) Communauté de père ; 2) Différence (spécification) par rapport au père ; 3) Communauté avec les concepts frères (situés au même niveau) ; 4) Différence par rapport aux concepts frères (sinon il n’aurait pas lieu de le définir). Ces principes correspondent à l’engagement sémantique (Bachimont, 2000) qui assure que chaque concept aura un sens univoque et non contextuel associé. Deux concepts sémantiques sont identiques



si l’interprétation du terme/libellé à travers les quatre principes différentiels aboutit à un sens équivalent.

- ❖ Ontologie de niveau 2 - Référentiel : Outre les caractéristiques énoncées au niveau précédent, les concepts référentiels (ou formels) se caractérisent par un terme/libellé dont la sémantique est définie par une extension d’objets. L’engagement ontologique spécifie les objets du domaine qui peuvent être associés au concept, conformément à sa signification formelle. Deux concepts formels seront identiques s’ils possèdent la même extension (par exemple, les « concepts » d’étoile du matin et d’étoile du soir associés à Vénus<sup>8</sup> « Instance »).
- ❖ Ontologie de niveau 3 - Opérationnel : Outre les caractéristiques énoncées au niveau précédent, les concepts du niveau opérationnel ou computationnel sont caractérisés par les opérations qu’il est possible de leur appliquer pour générer des inférences (engagement computationnel). Deux concepts opérationnels sont identiques s’ils possèdent le même potentiel d’inférence.

Les connaissances traduites par une ontologie sont véhiculées à l’aide des éléments suivants (Gomez-Perez, 1999) : 1) Concepts; 2) Relations ; 3) Fonctions ; 4) Attributs ; 5) Axiomes; 6) Instances.

### 3.3.1.3 Les concepts

Les connaissances modélisées dans une ontologie portent sur des objets auxquels on se réfère à travers des concepts (ou classes de l’ontologie) (Uschold & King, 1995). Le concept représente un ensemble d’objets, d’êtres, et leurs propriétés communes. Les concepts correspondent aux abstractions pertinentes d’un segment de la réalité (le domaine du problème, retenues en fonction des objectifs fixés et de l’application envisagée pour l’ontologie.

La notion, également appelée « intension » du concept (concept sémantique), est l’ensemble des caractéristiques que ces choses sont supposées avoir en commun. Elle est exprimée en termes de propriétés et d’attributs, de règles et de contraintes. Par exemple pour le concept « voiture », l’intension inclut les caractéristiques de « véhicule routier à moteur, généralement à 4 roues et 4 sièges pour une voiture à 4 personnes».

L’ensemble d’objets, également appelé « extension » du concept est la collection d’objets auxquelles la notion s’applique. Par exemple pour le concept « voiture », l’extension inclut « la Mazda MX5 avec le numéro d’immatriculation 2657ZB35 », « la voiture verte garée au coin de la rue en face de mon bureau ». Un concept est ainsi doté d’une sémantique référentielle (celle imposée par son extension) et d’une sémantique différentielle (celle imposée par son intension). Les concepts manipulés dans un domaine de connaissances sont organisés au sein d’un réseau de concepts. Dans une ontologie, l’ensemble des concepts est structuré hiérarchiquement et les concepts sont liés par des propriétés conceptuelles. La propriété utilisée pour structurer la hiérarchie des concepts est la subsomption. Elle lie deux

---

<sup>8</sup> Elle suit ou précède le soleil dans sa course dans le ciel, c’est pour cela qu’elle est visible tantôt en début de soirée, tantôt en fin de nuit.

concepts de la manière suivante : un concept C1 subsume un concept C2 si toute propriété sémantique de C1 est aussi une propriété sémantique de C2, c’est-à-dire si C2 est plus spécifique que C1. L’extension d’un concept est forcément plus réduite que celle d’un concept qui le subsume. Son intension est par contre plus riche. Une liste des principales propriétés pouvant être associées à un concept est donnée ci-dessous. La propriété portant sur un concept est :

- ❖ la **généricité** : un concept est générique s’il possède, dans une hiérarchie de concepts, la plus petite intension. Par exemple, le concept de « Maladie » est générique. En revanche, un concept générique possède une plus grande extension que ses sous-concepts.

Les propriétés portant sur deux concepts sont :

- ❖ l’**équivalence** : deux concepts sont équivalents s’ils ont une même extension. Par exemple, les concepts étoile du matin et étoile du soir associés à l’extension « Vénus » (3.3.1.2).
- ❖ la **subsomption** : un concept C1 subsume un concept C2 si toute propriété sémantique de C1 est aussi une propriété sémantique de C2, c’est-à-dire si C2 est plus spécifique que C1. Par exemple, le concept de « Maladie » subsume le concept de « Maladie Infectieuse ».
- ❖ l’**Incompatibilité** : deux concepts sont incompatibles si leurs extensions sont disjointes. Par exemple, homme et femme.

### 3.3.1.4 Les relations

Si certains liens conceptuels existant entre les concepts peuvent s’exprimer à l’aide de propriétés portées par les concepts (relation taxonomique), d’autres doivent être représentés à l’aide de relations autonomes. Une « Relation » est une notion d’association ou de lien s’établissant entre des concepts, exprimée en général par un terme ou par un symbole littéral général ou autre (ou plus généralement par un signe). Une relation permet de lier des instances de concepts, des instances de concepts aux valeurs (numérique, chaîne de caractères, etc.) ou des concepts génériques. Les relations traduisent les associations (pertinentes) existant entre les concepts présents dans le segment analysé de la réalité. Ces relations incluent les associations suivantes : (a) Sous-classe-de (généralisation-spécialisation) ; (b) Partie-de (agrégation ou composition) ; (c) Associée-à ; (d) Instance-de, etc. Ces relations nous permettent d’apercevoir la structuration et l’interrelation des concepts, les uns par rapport aux autres. Les propriétés intrinsèques à une relation sont :

- ❖ les propriétés algébriques :
  - symétrie : relation définie dans un ensemble E telle que, pour tout couple d’éléments (x, y) de  $E \times E$ , si x est en relation avec y, alors y est en relation avec x. par exemple, Paul « aime » Jeanne, Jeanne « aime » Paul.
  - réflexivité : relation pour laquelle pour tout a de l’ensemble X, a est relié à lui-même.
  - transitivité : relation pour laquelle une suite d’objets reliés consécutivement aboutit à une relation entre le premier et le dernier. Par exemple, Paul « est



l’ancêtre » de Jacques, Jacques « est l’ancêtre » de Marc, Paul « est l’ancêtre » de Marc.

- fonctionnelle : pour tout élément  $x$  de  $E$ ,  $x$  n’est en relation qu’avec 0 ou 1 élément  $y$  de  $F$ . Paul « a pour mère » Marie.
- ❖ la cardinalité : nombre possible de relations de ce type entre les mêmes concepts (ou instances de concept). Les relations portant une cardinalité représentent souvent des attributs. Par exemple, une pièce a au moins une porte.

Les propriétés liant deux relations sont :

- ❖ l’incompatibilité : deux relations sont incompatibles si elles ne peuvent lier les mêmes instances de concepts. Par exemple, les relations « a pour symptôme » et « a pour site anatomique » sont incompatibles.
- ❖ l’inverse : deux relations binaires sont inverses l’une de l’autre si, quand l’une lie deux instances  $I1$  et  $I2$ , l’autre lie  $I2$  et  $I1$ . Par exemple, les relations « a pour parent » et « a pour enfant » sont inverses l’une de l’autre.
- ❖ l’exclusivité : deux relations sont exclusives si, quand l’une lie des instances de concepts, l’autre ne lient pas ces instances, et vice-versa. L’exclusivité entraîne l’incompatibilité. Par exemple, l’appartenance et la non-appartenance sont exclusives.

La propriété liant une relation et des concepts est :

- ❖ le lien relationnel (propriété proposée par G. Kassel) : il existe un lien relationnel entre une relation  $R$  et deux concepts  $C1$  et  $C2$  si, pour tout couple d’instances des concepts  $C1$  et  $C2$ , il existe une relation de type  $R$  qui lie les deux instances de  $C1$  et  $C2$ . Un lien relationnel peut en outre être contraint par une propriété de cardinalité, ou porter directement sur une instance de concept (Kassel, 2002). Par exemple, il existe un lien relationnel entre les concepts « texte » et « auteur » d’une part et la relation « a pour auteur » d’autre part. on parle aussi de domaine et de co-domaine.

### 3.3.1.5 Fonctions

Les fonctions constituent des cas particuliers de relations, dans lesquelles un élément de la relation, le nième (extrant) est défini en fonction des  $n-1$  éléments précédents (intrants). Formellement une fonction  $F$  est définie par :

$$F: C1 * C2 * C3 * ... ..... Cn - 1 \rightarrow Cn$$

### 3.3.1.6 Attributs

Les attributs décrivent les propriétés des instances d’une classe. Ils sont toujours exprimés en termes de type de données ou domaine.

### 3.3.1.7 Axiomes

Les axiomes constituent des assertions, acceptées comme vraies, à propos des abstractions du domaine traduites par l’ontologie. Les éléments (concepts, relations, valeur, etc.) de l’ontologie doivent vérifier ses conditions afin de garantir la cohérence et permettre d’inférer de nouvelles connaissances.

### 3.3.1.8 Instances

Les instances constituent la définition extensionnelle de l’ontologie. Ces objets véhiculent les connaissances (statiques, factuelles) à propos du domaine.

### 3.3.1.9 Les règles

Les règles permettent d’enrichir les connaissances de l’ontologie. Ces règles capturent les dépendances entre les propriétés définies dans l’ontologie.

### 3.3.1.10 Conclusion

Les ontologies constituent un outil adéquat pour la représentation, le partage et la réutilisation des connaissances d’un domaine. Cependant, pour mettre en œuvre les ontologies afin qu’elles soient exploitables par les systèmes informatiques, des langages de représentation des connaissances doivent être utilisés. Un des langages les plus connus et les plus utilisés par les acteurs du web sémantique est le « *Web Ontology Language* » (OWL) (Cardoso, 2007; McGuinness & Harmelen, 2004) qui implémente les logiques de description (LD) (Baader, 2003).

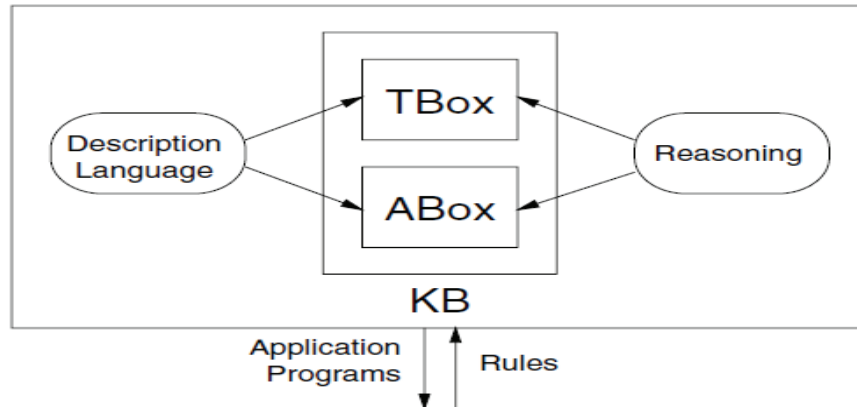
## 3.3.2 Les langages de représentation des connaissances

### 3.3.2.1 Les logiques de description

Les LD sont une famille de langages de représentation de connaissance qui peuvent être utilisés pour représenter les connaissances terminologiques d’un domaine d’application d’une manière formelle et structurée. Le nom de « logique de description » se rapporte, d’une part à la description des concepts utilisés pour décrire un domaine et d’autre part à la sémantique fondée sur la logique qui peut être donnée par une transcription en logique des prédicats du premier ordre (LPO). Les LD exploitent en général un sous-ensemble décidable de la LPO. Un problème de raisonnement est décidable s’il existe un algorithme permettant en un nombre fini d’étapes de dire s’il est vrai ou faux.

Les LD ont été développées comme une extension des *frames* et des réseaux sémantiques qui ne possédaient pas de sémantique formelle fondée sur la logique. Dans les logiques de description la connaissance est divisée en deux parties (Baader, 2003) (Figure 2):

- ❖ les informations terminologiques (TBox) qui sont des définitions de notions basiques. Ces informations sont « génériques » ou « globales », vraies dans tous les modèles et pour tous les individus.
- ❖ les informations sur les individus (ABox) sont « spécifiques » ou « locales », vraies pour certains individus particuliers.



**Figure 2: Architecture des systèmes de représentation des connaissances s'appuyant sur les logiques de description** (Baader, 2003).

Les éléments de base des logiques de description sont :

- ❖ les concepts qui correspondent à des classes d'éléments et qui dénotent un ensemble d'individus. Par exemple, on peut avoir le concept « Patient » qui dénote l'ensemble des patients suivants {Jean, Paul, Jacques, Georges, etc.}.
- ❖ les relations qui correspondent aux liens entre les éléments et sont interprétées comme des relations binaires sur un univers donné. En plus des concepts et des relations atomiques, toutes les logiques de description donnent la possibilité de construire des descriptions complexes de concepts et de relations.

Les systèmes de représentation des logiques de description n'enregistrent pas seulement les TBox et ABox, ils fournissent aussi des services pour faire des inférences sur ces descriptions. Les tâches de raisonnements sont les suivantes la subsumption, la satisfiabilité, la cohérence et la vérification d'instance.

On distingue classiquement les raisonnements qui portent sur la TBox et ceux portant sur la ABox. Les raisonnements qui portent sur la TBox sont les suivants (Baader, 2003):

- ❖ Le test de subsumption qui consiste à vérifier si une classe ou une propriété en subsume une autre. Le test de subsumption est à la base des opérations dites de classification qui consistent à déterminer la place d'une classe dans les hiérarchies de classes ou d'une propriété dans la hiérarchie des propriétés, c'est-à-dire déterminer les ascendants et descendants directs d'une classe ou d'une propriétés. Pour ces tests, il existe deux types d'algorithmes :
  - ceux fondés sur la normalisation des descriptions et leur comparaison. Les descriptions de classe et des propriétés sont normalisées en remplaçant récursivement les classes et les propriétés définies qui y interviennent par leur définition. Ce sont les formes normales des descriptions de classe ou de propriétés qui ne comportent plus que des classes et des propriétés primitives qui sont alors comparées. Et,
  - ceux fondés sur la méthode des tableaux sémantiques. Un test de subsumption entre deux classes ou propriétés est reformulé en test de satisfiabilité de la conjonction d'une description et de la négation de l'autre. Ce test de satisfiabilité est résolu par la méthode des tableaux sémantiques.

- ❖ Le test de satisfiabilité d’une classe consiste à vérifier qu’une classe admet des instances. En d’autres termes, il consiste à vérifier qu’il n’y pas de contradiction.
- ❖ Le test de cohérence d’une ABox permet de découvrir si son ensemble d’affirmations est cohérent, c’est-à-dire s’il a un modèle.
- ❖ Le test d’instanciation consiste à vérifier si un individu est instance d’une classe. Pour cela, on recherche les classes les plus spécifiques dont un individu est instance. Cette opération peut se ramener à l’opération de classification décrite ci-dessus en considérant une définition de classe abstraite de la description de l’individu considéré.

Dans un système de représentation fondé sur les logiques de description toutes les informations connues sont modélisées comme un couple  $(T, A)$ , où  $T$  est un ensemble de formules relatives aux informations terminologiques (la TBox) et  $A$  est un ensemble de formules relatives aux informations sur les assertions (la ABox). Il existe plusieurs types de logiques de description classés en fonction de leur expressivité. L’une des premières logiques de description est le langage  $\mathcal{FL}^-$ , qui est défini comme une logique de description permettant l’utilisation des quantificateurs universels, de la conjonction, et des quantificateurs existentiels de la forme  $\exists R.T$ . Le langage  $\mathcal{FL}^-$  a été proposé comme un formalisme pour la sémantique des *frames* de Minsky (Minsky, 1975). La conjonction de concepts est implicite dans la structure d’un *frame*, qui requiert un ensemble de conditions pour être satisfait. La quantification des rôles permet de caractériser les slots. La logique  $\mathcal{AL}$  « *Attributive Language* », a étendu la logique  $\mathcal{FL}^-$  en y ajoutant la négation des concepts atomiques. Cette logique peut être considérée comme la logique de base des autres logiques de description. Les logiques de description qui existent sont des combinaisons des différents éléments, par exemple, si on rajoute la négation complète  $\mathcal{C}$  à la logique  $\mathcal{AL}$ , on obtient la logique  $\mathcal{ALC}$ . Certaines logiques sont équivalentes, notamment  $\mathcal{ALC}$  et  $\mathcal{ALUE}$ . Ces deux logiques augmentées par  $\mathcal{R}^+$  sont notées  $\mathcal{S}$ . Les langages utilisés par OWL en sont des extensions, respectivement  $\mathcal{SHIF}$ ,  $\mathcal{SHIF}^{(D)}$  pour OWL-Lite,  $\mathcal{SHOIN}$ ,  $\mathcal{SHOIN}^{(D)}$  pour OWL-DL,  $\mathcal{SROIQ}^{(D)}$  pour OWL 2 (Cuenca et al., 2008).

### 3.3.2.2 Web Ontology Language : OWL

OWL est le résultat des efforts du *World Wide Web Consortium* (W3C) pour créer un langage standard pour le web sémantique. Il est fondé sur un sous-ensemble des LD. OWL est le langage de représentation des ontologies le plus utilisé (Cardoso, 2007; Dameron, Rubin, & Musen, 2005). De nombreuses conversions des terminologies existantes en OWL ont été faites (par exemple, MeSH (Soualmia, Golbreich, & Darmoni, 2004), FMA (Dameron et al., 2005) ou NCIT (Golbeck et al., 2003)). Le langage OWL se présente en deux grandes déclinaisons, OWL 1 et OWL 2.

OWL 1 a trois versions : OWL-L, OWL-DL, OWL-Full (Antoniou & Van Harmelen, 2001). Ces versions varient entre elles par leur degré d’expressivité sémantique et leur niveau d’exploitabilité informatique. OWL-Lite est le plus simple, il permet de représenter des hiérarchies. OWL-DL est plus expressif sur le plan sémantique mais pose des difficultés aux

niveaux de l’exploitation d’autant plus que les connaissances à représenter sont volumineuses. OWL-Full est le langage le plus expressif sur le plan sémantique, mais le moins décidable.

OWL 2 (Cuenca et al., 2008) étend l’expressivité de OWL 1 en ajoutant la prise en charge :

- ❖ de la négation des relations, par exemple, Jean « n’a pas » de douleur au dos.
- ❖ des restrictions avec des cardinalités qualifiées. Il est possible d’exprimer le fait qu’une maladie « peut avoir » n signes parmi m.
- ❖ du chaînage des relations, par exemple la relation « a pour oncle » qui découle de la conjonction des relations « a pour parent » et « a pour frère ». Une nouvelle relation peut être produite via d’autres relations.
- ❖ des restrictions réflexives, c’est-à-dire, une classe liée à elle-même via une relation. Par exemple, une personne narcissique est une personne qui « s’aime » soit même.
- ❖ des annotations, il est possible d’intégrer des annotations sur des axiomes, des classes, etc.

L’intérêt de l’utilisation du langage OWL repose sur le fait que :

- ❖ La majorité des ontologies existantes est représentée en OWL (Dameron et al., 2005; Golbeck et al., 2003; Soualmia et al., 2004).
- ❖ OWL est devenu le standard pour le Web sémantique recommandé par le W3C.
- ❖ La complexité du domaine médical impose d’avoir un langage qui a une grande expressivité (Antoniou & Van Harmelen, 2001; Wang et al., 2006).
- ❖ Des éditeurs d’ontologie puissants et conviviaux utilisant OWL comme langage de représentation sont disponibles, par exemple PROTÉGÉ (cf. 3.3.2.5).

### 3.3.2.3 Formalisme de règles

#### a. Rule Interchange Format (RIF)

RIF « *Rule Interchange Format* » (Gordon, Governatori, & Rotolo, 2009) est un formalisme proposé par le W3C en 2005 permettant de fournir une interopérabilité entre les langages de règles en général et ceux utilisés pour le web en particulier tels que le « *Semantic Web Rule Language* » (SWRL), RuleML, Common logic et « *Semantic Business Vocabulary Rule* » (SVBR).

Le but de RIF est de permettre la traduction des règles d’un langage à un autre, et ainsi de permettre les transferts des règles d’un formalisme à un autre. Le formalisme RIF a une architecture en couches qui renferme les composants suivants : RIF Core, RIF Basic Logic Dialect (RIF-BLD), RIF Production Rule Dialect (RIF-PRD) et RIF Framework for Logic Dialects (RIF-FLD). L’utilisation du formalisme RIF est limitée actuellement car il n’existe pas d’outils ou de *plug-ins* (Elenius, 2012) pour la gestion des règles RIF intégrés dans PROTÉGÉ. Quant à son intégration au formalisme OWL, des travaux sont en cours (Cuenca et al., 2008).

## b. Clause de Horn

Une clause de Horn (Horn, 1951) est une clause comportant au plus un littéral positif, par exemple,  $\neg p \vee \neg q \vee \dots \neg v \vee \neg t \dots \vee u$ . Un littéral est un atome ou une variable (aussi appelé littéral positif) ou la négation d'un atome ou d'une variable (aussi appelé littéral négatif).

Les clauses de Horn strictes sont équivalentes à des règles « si ... alors ... » et permettent de déduire de nouveaux faits à partir de faits existants. Un exemple de clause de Horn stricte, est  $\text{Onclede}(z,y) :- \text{estParentde}(x,y), \text{estFrerede}(z,x)$ . Le langage Prolog (Gacogne, 2009) s'appuie sur la théorie des clauses de Horn.

## c. Semantic Web Rule Language: SWRL

**SWRL**<sup>9</sup> « *Semantic Web Rule Language* » (I. Horrocks et al., 2004) est un langage de règles pour le web sémantique, combinant le langage OWL-DL et le langage RuleML « *Rule Markup Language (Unary/Binary Datalog)* ». Avec **SWRL**, les connaissances peuvent être exprimées sous la forme : antécédent  $\rightarrow$  conclusion. Les axiomes du langage OWL-DL sont étendus à l'aide des clauses de Horn (Horn, 1951) réduites aux prédicats unaires et binaires. Ce langage est indécidable, les règles SWRL sont appliquées même si les individus ne sont pas présents dans la base. Une restriction de SWRL, appelée DL-safe rules, a été conçue pour conserver la décidabilité, c'est-à-dire que les règles écrites ne portent que sur des individus explicitement présents dans la base. Autrement dit, les règles ne peuvent s'appliquer que si l'identité de toutes les instances présentes est connue. De nombreux moteurs d'inférences commencent à supporter SWRL : Bossam, Hoolet, KAON2, Pellet, RacerPro, R2ML « *REVERSE Rule Markup Language* » et Sesame. Ils suivent trois types d'approches :

- ❖ Traduire SWRL en logique du premier ordre (Hoolet) et appliquer un algorithme de chaînage avant.
- ❖ Traduire OWL-DL en règles et appliquer un algorithme de chaînage avant (Bossam).
- ❖ Intégrer les règles SWRL dans le moteur d'inférences OWL-DL fondé sur les algorithmes des tableaux sémantiques (Pellet) (Sirin, Parsia, Grau, Kalyanpur, & Katz, 2007).

L'intérêt de l'utilisation de SWRL repose sur le fait qu'il permet d'exprimer des connaissances qu'on ne peut pas exprimer en OWL (Javanmardi, Amini, & Jalili, 2006). Par exemple, avec OWL 1 on ne peut pas exprimer le fait que x est l'oncle de y parce que z est à la fois parent de x et le frère de y. En SWRL,  $\text{estParentde}(?z,?y) \wedge \text{estFrerede}(?x,?z) \rightarrow \text{estOnclede}(?x,?y)$ . En OWL 2, il n'est pas possible d'exprimer le fait que x est l'enfant d'un couple marié (y,z) car OWL 2 ne permet pas d'exprimer des relations entre des instances qui sont liées à une instance donnée. En SWRL,  $\text{Personne}(?x) \wedge \text{aPourPere}(?x,?y) \wedge \text{aPourMere}(?x,?z) \wedge \text{estMariee}(?y,?z) \rightarrow \text{estFilsduCouple}(?x)$ .

SWRL est aussi requis pour représenter des expressions contenant des opérateurs mathématiques (addition, quotient, soustraction, produit, etc.) (Prcela, Gamberger, & Jovic, 2008).

<sup>9</sup> <http://www.w3.org/Submission/SWRL/>



### 3.3.2.4 Ontology design Pattern (ODP)

#### a. Définitions

Un patron de conception d’ontologies « *Ontology Design Pattern* » (ODP<sup>10</sup>) est un modèle générique pour résoudre des problèmes de modélisation récurrents (Gangemi & Presutti, 2009). Cette notion tire ses origines des travaux de l’architecte et mathématicien Christopher Alexander qui introduit la notion de patron de conception « *design pattern* » (DP) comme un moyen de partage des bonnes pratiques pour la résolution des problèmes de modélisation. Cependant, c’est dans le domaine de l’ingénierie logicielle que les patrons de conception ont été largement utilisés depuis le milieu des années 90. Ils fournissent (Harrison, Avgeriou, & Zdun, 2007) :

- ❖ un vocabulaire commun pour la communication entre programmeurs.
- ❖ une documentation des solutions logicielles.
- ❖ des moyens pour la génération automatique du code informatique.

En effet, dans le domaine de l’ingénierie logicielle, les patrons de conception répondent à des problèmes de conception de logiciels dans le cadre de la programmation orientée objet. Ce sont des solutions connues et éprouvées dont la conception provient de l’expérience des programmeurs. D’une manière analogue à un patron de couture, le DP décrit les grandes lignes d’une solution, qui peuvent ensuite être modifiées et adaptées en fonction des besoins. Les DP décrivent des procédés de conception généraux et permettent en conséquence de capitaliser l’expérience appliquée à la conception de logiciel. Ils ont une influence sur l’architecture logicielle d’un système informatique. Il n’y a aucun aspect théorique dans les DP, notamment pas de formalisation (à la différence des algorithmes) (Laurent, 2007).

Tout comme dans le domaine de l’ingénierie logicielle, le domaine de l’ingénierie ontologique produit et manipule des modèles, et les ODP sont des modèles de solutions abstraites relatives à des problèmes clairement définis. Ces ODP sont fondés sur les meilleures approches de résolutions des problèmes de modélisation et elles facilitent la conception des ontologies (Gangemi, Gomez-Perez, Presutti, & Suarez-Figueroa, 2008).

#### b. Patrons de conception ontologiques existants

##### Ontology Design Pattern 1 (ODP 1)

Une des limites inhérentes de OWL, RDF et SWRL est qu’ils ne gèrent que les relations d’arité 1 ou d’arité 2 (Stevens et al., 2007). Pour exprimer une relation d’arité supérieure à 2 (ou relation d’arité n), nous devons utiliser l’ODP 1 du W3C (Hayes et al., 2006). L’ODP 1 exprime ce type de relations via un processus appelé « réification ». La réification dans ce contexte consiste à exprimer une relation d’arité supérieure à 2 par une classe.

---

<sup>10</sup> [http://ontologydesignpatterns.org/wiki/Main\\_Page](http://ontologydesignpatterns.org/wiki/Main_Page)

### Chaînage des relations (Property chain)

OWL 1 ne permet pas de gérer les chaînages des relations et certains chaînages des relations<sup>11</sup> ne sont pas gérables par OWL 2 (cf. 3.3.2.3) (Javanmardi et al., 2006). Le chaînage des relations consiste à faire une conjonction de relations pour en produire une nouvelle, par exemple, la relation « estOnclede » est le résultat du chaînage des relations « estParentde » et « estFrerede », par exemple,  $\text{estParentde}(?z,?y) \wedge \text{estFrerede}(?x,?z) \rightarrow \text{estOnclede}(?x,?y)$ . Un outil efficace pour la représentation de tous les types de chaînage des relations est le formalisme SWRL (cf. 3.3.2.3).

### Axiomes de fermeture (Closure axiom)

OWL est fondé sur l’hypothèse de monde ouvert (Open World Assumption). Cela signifie que tout ce qui n’est pas explicitement décrit est inconnu (et non pas faux comme lorsque l’on raisonne en monde fermé). Par exemple, si l’on veut décrire un carnivore comme quelque chose qui ne mange que de la viande, la relation existentielle ( $\exists \text{mange}(\text{viande})$ ) est insuffisante. La relation existentielle signifie qu’un carnivore doit manger de la viande mais peut manger autre chose. Pour préciser qu’il ne doit manger que de la viande, il faut « fermer » la relation (par un axiome de fermeture<sup>12</sup>) en ajoutant une restriction munie d’une relation avec un quantificateur universelle :  $\forall \text{mange}(\text{viande})$ . Elle signifie que la classe « carnivore » (ou ses sous-classes ou ses instances) ne peut avoir de relations « mange » qu’avec la classe « viande » (ou ses sous-classes ou ses instances) (Horridge, Knublauch, Rector, Stevens, & Wroe, 2004). L’axiome de fermeture, en explicitant complètement l’ensemble d’arrivée de la relation permet de se rapprocher d’un raisonnement en monde fermé.

#### 3.3.2.5 PROTÉGÉ : un éditeur d’ontologies

PROTÉGÉ<sup>13</sup> est un environnement graphique de création et de manipulation des ontologies développé par le « *Stanford Medical Informatics group* » (SMI). Il est l’éditeur d’ontologies dominant (Malviya, Mishra, & Sahu, 2011) toutes disciplines confondues avec une communauté de quelques 50 000 utilisateurs représentant des projets industriels et de recherche dans plus de 100 pays (Cardoso, 2007; Malviya et al., 2011). Il permet d’éditer, de visualiser, de contrôler (vérification des contraintes) les ontologies. Il facilite l’extraction d’ontologies à partir de sources textuelles et fournit des services inférentiels (cf. 3.3.2.1). D’autre part PROTÉGÉ permet d’exporter une ontologie vers des formats variés tels que RDF/S, OWL et schémas XML. Fondé sur une architecture ouverte, PROTÉGÉ peut être complété par de nouvelles fonctionnalités grâce à l’ajout de nouveaux composants (interfaces : *plug-ins*) dont certains sont élaborés par les utilisateurs. Plus de 70 *plug-ins* fournissent des services tels que la visualisation d’ontologies (OntoViz), l’alignement d’ontologies (PROMPT) et des interfaces avec des moteurs de règles (par exemple, Jess<sup>14</sup>) et des formalismes de règles (par exemple, SWRL) (Bodenreider & Stevens, 2006).

<sup>11</sup> <http://www.gong.manchester.ac.uk/odp/html/CompositePropertyChain.html>

<sup>12</sup> <http://www.gong.manchester.ac.uk/odp/html/Closure.html>

<sup>13</sup> <http://protege.stanford.edu>

<sup>14</sup> <http://www.jessrules.com/jess/index.shtml>



Les éléments de base du langage OWL pris en charge par PROTÉGÉ sont (Horridge et al., 2004; Malviya et al., 2011):

- ❖ Les individus ou instances : Un individu est une instance de classe. Par exemple, on pourrait dire que « Jean » est une instance de la classe « Patient ».
- ❖ Les propriétés ou relations (ObjectProperty, DatatypeProperty) ou slots dans le cas des *frames* (Minsky, 1975): Une propriété est une relation binaire entre deux instances, elle lie deux instances. Par exemple, on pourrait dire que « Paul » et « Jeanne », deux instances de la classe « Humain » s’aiment. Une propriété peut être symétrique, c’est le cas de « aime », deux propriétés peuvent être inverses l’une de l’autre, c’est le cas de « aPourFils » et « estFilsde ». Une propriété peut être fonctionnelle, c’est-à-dire n’avoir qu’une seule valeur, c’est le cas de « aPourAge ». Les propriétés de type « *ObjectProperty* » permettent de lier deux instances de classes, par exemple **Paul** « estFrèrede » **Jacques**. En revanche les propriétés de type « *DatatypeProperty* » permettent de lier une instance à une valeur, par exemple **Paul** « aPourAge » **18**.
- ❖ Les classes (Class) ou concepts : Une classe est un ensemble qui contient des instances. Elle est décrite via des critères ou des conditions (encore appelés des restrictions) formels qui définissent les pré-requis pour être membre de cette classe. Par exemple, on peut dire que pour être étudiant à l’université de Rennes il faut être détenteur d’une carte d’étudiant de cette université. Les classes peuvent être organisées en hiérarchie de classe (superclasse et sous-classe), encore appelée taxonomie. Une sous-classe est subsumée par sa superclasse. Par exemple, la classe « Étudiant boursier » est une sous-classe de la classe « Étudiant ». Deux types de classes existent dans PROTÉGÉ, les classes dites primitives et les classes dites définies. Les classes dites primitives sont celles qui ne sont introduites que par des conditions nécessaires, et les classes dites définies sont celles qui sont introduites par des conditions nécessaires et suffisantes. Une condition nécessaire est celle que toutes les instances d’une classe doivent avoir. Par exemple, si on est atteint d’un myélome multiple (MM) alors on a une quantité d’immunoglobuline de type A  $\geq 3\text{g/dL}$ . En revanche, pour être diagnostiqué comme atteint d’un MM un patient doit avoir une quantité d’immunoglobuline de type A  $\geq 3\text{g/dL}$ , une présence de plasmocyte dans la moelle osseuse  $\geq 10\%$  et une présence de CRAB (hypercalcémie, insuffisance rénale, anémie et lésion osseuse). Ces conditions sont mises en œuvres via des restrictions. Une restriction décrit les conditions que les instances doivent remplir pour être membre d’une classe donnée. Par exemple, on pourrait dire qu’un patient atteint d’une rougeole doit avoir le signe de Koplik. Le signe de Koplik étant une condition nécessaire et suffisante pour être diagnostiqué comme étant atteint de la rougeole.

### 3.3.2.6 Editeur de règles : SWRLTab

SWRLTab est essentielle pour l’édition des règles SWRL. L’éditeur de règles dans PROTÉGÉ offre un assistant à la saisie et une liste d’objets permettant de facilement écrire des règles (Figure 3).

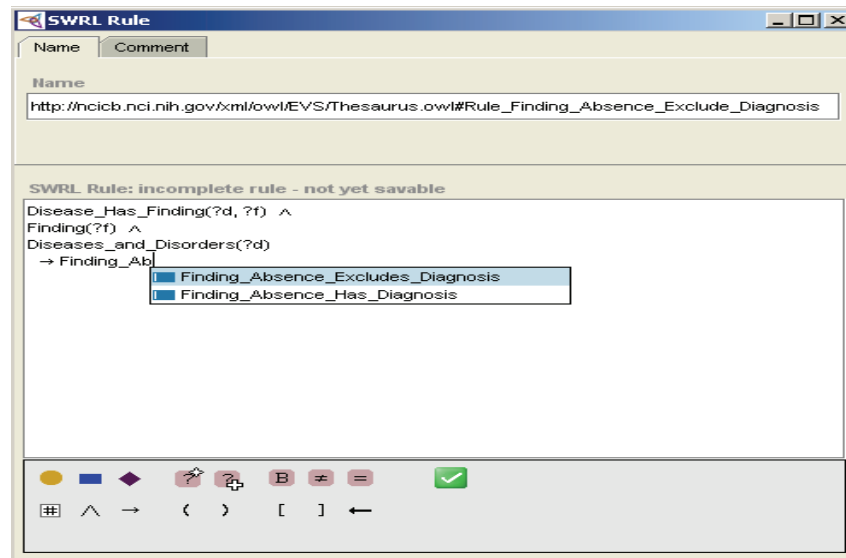


Figure 3 : Écran de saisie d’une règle SWRL sous PROTÉGÉ 3.4.8. L’éditeur offre un système d’auto-complétion pour faciliter la saisie des règles.

### 3.3.2.7 Moteur d’inférences

Pour exploiter une ontologie et les règles, un ou des moteurs d’inférences sont requis. De nombreux moteurs d’inférences existent tels que Pellet (Sirin et al., 2007), OWLIM (Bishop et al., 2011; Kiryakov, Ognyanov, & Manov, 2005), KAON (Motik & Sattler, 2006), Hermit (Shearer, Motik, & Horrocks, 2008), FACT++ « *Fast Classification of Terminologies* » (Tsarkov & Horrocks, 2006), Racer « *Renamed ABox and Concept Expression Reasoner* » (Haarslev & Ralf, 2000), pour ne citer que ceux-là. Ces moteurs d’inférences sont tous déductifs<sup>15</sup>. Ils permettent de :

- ❖ vérifier la cohérence pour s’assurer que l’ontologie ne contienne pas de fait contradictoires. En logique de description, cette opération vérifie si la ABox est conforme à la TBox.
- ❖ vérifier la satisfiabilité des concepts pour détecter les classes qui peuvent ne pas avoir d’instances.
- ❖ faire de la classification qui permet de dire qu’une classe donnée est sous-classe d’une autre. Cette opération crée une hiérarchie de classes et elle permet de produire des réponses à d’autres types de requêtes comme : lister toutes les sous-classes directes d’une classe donnée.
- ❖ faire de la réalisation qui permet de trouver les classes les plus spécifiques à une instance.

### 3.3.2.8 Conclusion

Les ontologies permettent de représenter, de partager et de réutiliser les connaissances d’un domaine. L’avantage des ontologies est qu’elles permettent d’allier au modèle théorique des règles (règle de production, règle métier, etc.) (Bruaux, 2007; Page, 1990). Les formalismes de représentation OWL et SWRL offrent suffisamment d’outil pour mettre en œuvre les modèles de connaissances ontologiques pour des domaines complexes, et pour des cas

<sup>15</sup> C’est cette caractéristique qui fait que ces moteurs d’inférences n’arrivent pas à gérer les tâches de raisonnement liées au diagnostic médical car ce dernier est abductif (cf. 2.3.3 et 3.3.4.6).

particuliers (par exemple, relation d’arité  $n$ ) les ODP existantes (cf. 3.3.2.4) peuvent être utilisés.

Cependant, malgré ces attributs, selon Burgun et Bertaud-Gounot (Bertaud-Gounot, Duvauferrier, et al., 2012; Burgun, Bodenreider, & Jacquelinet, 2005) les ontologies classiques et les moteurs d’inférences classiques ne permettent pas de représenter et de raisonner sur les connaissances liées au raisonnement médical, les connaissances probabilistes, les connaissances imprécises et vagues, et les connaissances liées aux signes négatifs. En effet, ces ontologies n’intègrent que les connaissances sur les maladies (processus pathologique, localisation, etc.), et non les connaissances (critères diagnostiques) en rapport avec le processus du diagnostic médical (cf. 2.3). Pour ce qui est des moteurs d’inférences classiques (Pellet, OWLIM, KAON, Hermit, FACT++, Racer), ils sont tous déductifs et ne peuvent prendre en charge le raisonnement médical, car ce dernier est abductif (Josephson et al., 1996; Walton, 2004) (cf. 2.3.3).

### 3.3.3 Système informatisé d’aide au diagnostic en urgences odontologiques

Ce système s’appuie sur une base de connaissances ontologiques contenant les descriptions des pathologies rencontrées en urgences odontologiques ainsi que les critères diagnostiques (signes, symptômes, les tests diagnostiques) associés (Bertaud-Gounot et al., 2010; Le Moing, 2009). L’éditeur d’ontologie utilisé est PROTÉGÉ 3.4.2 et le moteur d’inférences utilisé est Pellet 1.4.2. L’interface d’utilisation du système est celui de PROTÉGÉ 3.4.2.

L’évaluation du système sur la base de 250 dossiers a permis de ressortir les limites suivantes (Le Moing, 2009): (a) La première limite repose sur le fait que les moteurs d’inférences classiques du web sémantique sont tous déductifs (cf. 3.3.2.7). Ils ne permettent donc pas de faire du diagnostic médical (cf. 2.3) car ce dernier est abductif (cf. 2.3.3). Faire de la classification diagnostique avec ces moteurs d’inférences se résume à faire de la subsomption. Pour que ces moteurs d’inférences nous donnent des résultats corrects (la maladie dont souffre le patient), il faut leur transmettre au minimum tous les paramètres (signes, symptômes) tels qu’ils se trouvent dans la description des maladies, ce qui dans la pratique cause d’énormes problèmes car tous les signes ne sont pas souvent observés. Par exemple, en reprenant le cas du myélome multiple (MM) (cf. 3.3.2.5), un patient doit avoir une quantité d’immunoglobuline de type A  $\geq 3\text{g/dL}$ , une présence de plasmocyte dans la moelle osseuse  $\geq 10\%$  et une présence de CRAB (hypercalcémie, insuffisance rénale, anémie et lésion osseuse). Si on observe uniquement une quantité d’immunoglobuline de type A  $\geq 3\text{g/dL}$  et une présence de plasmocyte dans la moelle osseuse  $\geq 10\%$ , le système n’est pas capable de proposer le MM comme hypothèse diagnostique. Il faut observer tous les signes décrivant le MM au minimum. (b) La deuxième limite concerne le fait que le langage OWL (cf. 3.3.2.2 et 3.3.2.4) et les moteurs d’inférences associés ne prennent pas en charge de manière native les probabilités (les sensibilités et spécificités des signes pour les maladies, et les prévalences des maladies pour une population donnée). En effet, l’expression des connaissances probabilistes fait intervenir des relations ayant des arités supérieures à 2, ce qu’OWL ne permet pas d’exprimer (Hayes et al., 2006; Stevens et al., 2007) (cf. 3.3.2.4). (c)

La troisième et dernière limite est l’interface (celle de PROTÉGÉ) d’utilisation non conviviale qui rend l’exploitation dudit système difficile.

### 3.3.4 Ontology-driven differential diagnosis based on logical inference and probabilistic refinements (ODDIN)

ODDIN (García-crespo et al., 2010) est un système d’aide au diagnostic médical dans le domaine de la médecine interne. Ce système s’appuie sur :

- ❖ une ontologie ad hoc (Rodríguez-Gonzalez et al., 2010) faite en OWL-DL contenant les descriptions des maladies et les critères diagnostiques (signes, symptômes, les tests). Cette ontologie contient aussi des critères diagnostiques faisant références à d’autres maladies, d’où la notion de maladie multi-niveau ou diagnostic multi-niveau « *multi-level diagnosis* » (DMN) (Figure 4) (Rodríguez-González et al., 2011; Rodríguez-González, Labra-Gayo, et al., 2012).
- ❖ un ensemble de règles développé avec le langage « JenaRules » (The Apache Software Foundation, 2008). Ces règles sont stockées dans un fichier différent de celui de l’ontologie. JenaRules est fondé sur le langage RDF(S) (RDF Working Group, 2004) et il utilise une représentation sous forme de triplets RDF.
- ❖ un fichier externe contenant un ensemble de probabilités *a priori* ou prévalences et des probabilités conditionnelles (sensibilité et spécificité des signes pour les maladies). Et,
- ❖ un moteur d’inférences déductif et probabiliste. Il utilise le moteur d’inférences Jena pour gérer les règles écrites en JenaRules, le moteur d’inférences Pellet pour l’ontologie et une inférence bayésienne (Smets, 2005) pour trier (classer, ordonner) les diagnostics produits.

Les résultats fournis lors des premières phases d’évaluation du système ODDIN sont satisfaisants à hauteur de 86,7 % de taux de diagnostics corrects (le premier diagnostic dans la liste des hypothèses de maladies (diagnostic) produites par ODDIN) sur la base de 336 cas observés (García-crespo et al., 2010; Rodríguez-González, Torres-Niño, et al., 2012).

Cependant, ODDIN présente certaines limites : (a) son moteur d’inférences ne gère pas les DMN, (b) son modèle de représentation ontologique n’intègre pas les signes absents et (c) les probabilités ne sont pas représentées dans l’ontologie.

Des propositions ont été faites par Rodríguez-González (Rodríguez-González & Alor-Hernández, 2013; Rodríguez-González et al., 2011; Rodríguez-González, Labra-Gayo, et al., 2012) pour la gestion des DMN.

#### 3.3.4.1 Ontologie et définitions des maladies multi-niveaux

Le DMN concerne de manière générale les cas où une maladie est un signe pour une autre maladie. Par exemple, le myélome multiple (MM) (Figure 25)<sup>16</sup> est caractérisé par une quantité d’immunoglobuline de type A  $\geq 3\text{g/dL}$ , une présence de plasmocyte dans la moelle osseuse  $\geq 10\%$  et une présence de CRAB (hypercalcémie, insuffisance rénale, anémie et lésion osseuse), et l’hypercalcémie est caractérisée par un taux de calcium sérique  $\geq 2,6$

<sup>16</sup> Une figure présentant concrètement un cas de représentation de diagnostic multi-niveau pour le myélome multiple.

mmol/L. Ce type de définition est illustré de manière générique par les figures suivantes. La prise en charge de ce type de représentation par les moteurs d’inférences du web sémantique pose certaines difficultés que nous allons présenter.

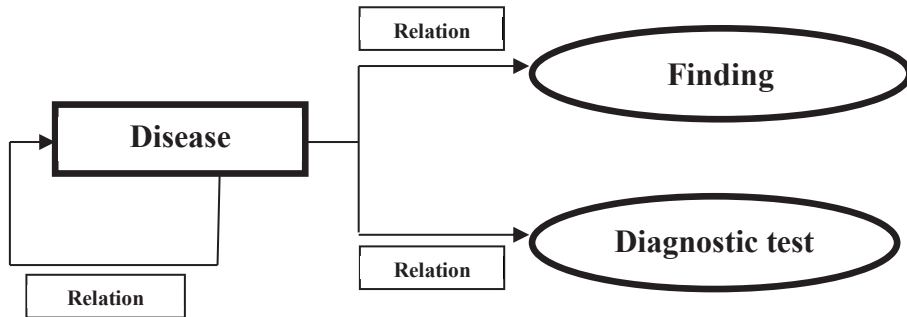


Figure 4: Description d’une maladie sous forme multi-niveau (DMN).

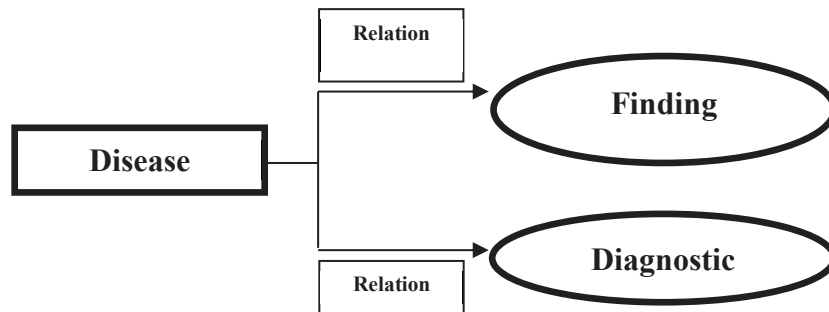


Figure 5: Description d’une maladie dite simple.

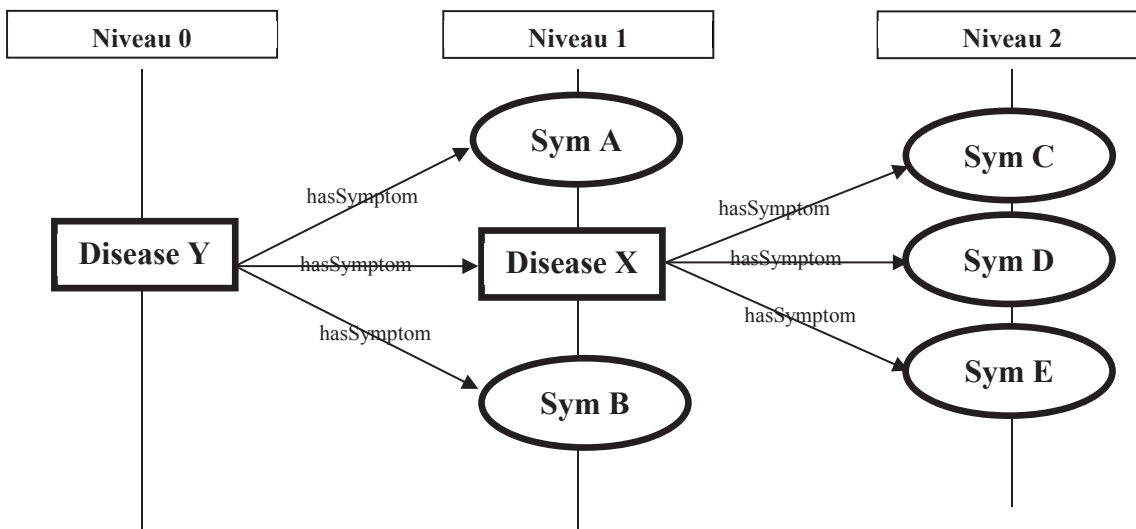


Figure 6 : Définition multi-niveau d'une maladie. « Disease Y » a une profondeur 2.

La gestion des diagnostics multi-niveaux s’appuie sur la propagation transitive (Seidenberg & Rector, 2006). Horrocks et Patel-Schneider ont fait ressortir le fait que les moteurs d’inférences implémentant les logiques de description classiques ne permettent pas de faire ce type d’inférences (propagation transitive) (I. Horrocks & Patel-schneider, 2004). Par exemple, soit le patient *p* ayant les signes et symptômes suivants : « Sym C, Sym D, Sym E » (Req 1) pour le premier cas, « Sym A, Sym B, Sym C, Sym D, Sym E » (Req 2) pour le second. Avec le moteur d’inférences « *Pellet* » (Sirin et al., 2007), la Req 1 infère la Disease X, et pour la Req 2 le système infère aussi la Disease X parce qu’elle est plus générale. Pellet

implémente les logiques de description classiques et il ne peut donner de réponses adéquates que si la Req 2 a la même description que les classes décrites dans l’ontologie, c’est-à-dire sous la forme suivante : « Sym A, Sym B, Disease X ». Si tel est le cas alors le moteur d’inférences infère la Disease Y.

Trois solutions ont été proposées par Rodríguez-González (Rodríguez-González & Alor-Hernández, 2013; Rodríguez-González et al., 2011; Rodríguez-González, Labra-Gayo, et al., 2012) : (a) l’approche « *ADONIS* » (Rodríguez-González et al., 2011), (b) l’approche « *SeDeLo* » (Rodríguez-González, Labra-Gayo, et al., 2012) et (c) une troisième approche (Rodríguez-González & Alor-Hernández, 2013) consistant à modifier les définitions des maladies au niveau de l’ontologie.

La première approche « *ADONIS* » (Rodríguez-González et al., 2011) consiste à créer de nouvelles classes sur la base des définitions des classes Disease X et Disease Y. Cette approche respecte l’hypothèse du monde fermé (cf. 3.3.2 et 3.3.2.3) et est implémentée en OWL 2 (Patel-Schneider, 2009). Premièrement, une classe de type énumération est créée :  $SymptomsX = oneOf \{SymC, SymD, SymE\}$ . Dans cette classe, les éléments  $\{SymC, SymD, SymE\}$  sont des instances des classes SymC, SymD et SymE. Ensuite, la classe « Disease X » est modifiée comme suit :

$$\begin{aligned} Disease\ X \equiv & \exists hasSymptom\ SymptomsX \wedge \\ & \forall hasSymptom\ SymptomsX \wedge \\ & hasSymptoms = 3 \end{aligned}$$

Cette nouvelle définition de « Disease X » stipule que ses signes sont dans l’ensemble  $\{SymC, SymD, SymE\}$  et uniquement dans cet ensemble (clôture d’axiomes), et que leur nombre ne doit pas excéder 3 (cardinalité qualifiée (Rector & Schreiber, 2005)). En appliquant la même procédure pour la « Disease Y », on obtient :  $SymptomsY = oneOf \{SymA, SymB\} \cup SymptomsX$  et,

$$\begin{aligned} Disease\ Y \equiv & \exists hasSymptom\ SymptomsY \wedge \\ & \forall hasSymptom\ SymptomsY \wedge \\ & hasSymptoms = 5 \end{aligned}$$

A partir de cette approche (Rodríguez-González et al., 2011) il est possible d’inférer la « Disease Y » pour notre requête « Req 2 ». L’utilisation des énumérations (*ou inclusif*) donne plus de souplesse aux inférences, ainsi quel que soit le nombre de signes observés, le système produira des résultats adéquats.

La deuxième approche « *SeDeLo* » (Rodríguez-González, Labra-Gayo, et al., 2012) consiste à étendre les fonctionnalités de ADONIS (Rodríguez-González et al., 2011). ADONIS, malgré sa capacité à gérer les diagnostics multi-niveaux ne permet pas de gérer les définitions de maladies dites simples (Figure 5) d’une part (Rodríguez-González, Labra-Gayo, et al., 2012). D’autre part, il ne permet pas d’inférer correctement lorsqu’on a des cas composés (Rodríguez-González, Labra-Gayo, et al., 2012). Considérons le patient p ayant la Disease X et, les signes SymA et SymB. Le système devrait exclure automatiquement la Disease X et inférer la Disease Y ou du moins proposer les deux cas en spécifiant à la limite le nombre de signe et les signes liés à chaque cas. Malheureusement, il ne le fait pas, il propose la Disease X. Pour pallier cette limite, l’approche SeDeLo ajoute des instructions au



niveau du moteur d’inférences permettant à ce dernier lors des cycles de raisonnement de : (a) mémoriser toutes les maladies liées aux signes observés et, (b) compter les signes observés et de les comparer avec le nombre de signes caractérisant chaque maladie mémorisée. Cela permet d’avoir un raisonnement incrémental permettant de présenter les différents résultats et les informations associées. Cependant, la durée des traitements est excessivement longue, car avec cette approche les temps d’inférence peuvent aller jusqu’à 9 heures pour une ontologie contenant 5 maladies et 11 signes (Rodríguez-González & Alor-Hernández, 2013).

La troisième approche (Rodríguez-González & Alor-Hernández, 2013) consiste à changer la structure de la Req 2 en mettant la Disease X dans la description de la Req 2 à la place de Sym C, Sym D et Sym E ou, à modifier la structure de l’ontologie initiale. Dans cette optique, toutes les définitions des maladies ayant pour signe d’autres maladies sont modifiées. Les maladies contenues dans ces définitions sont remplacées par leurs signes respectifs comme l’exemple ci-après pour le cas de Disease Y (Figure 7 et 8).

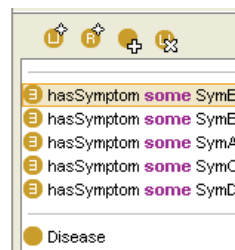


Figure 7: Nouvelle définition de la classe Disease Y. La classe Disease X a été éclatée en mettant les signes de cette dernière (Sym C, Sym D, Sym E).

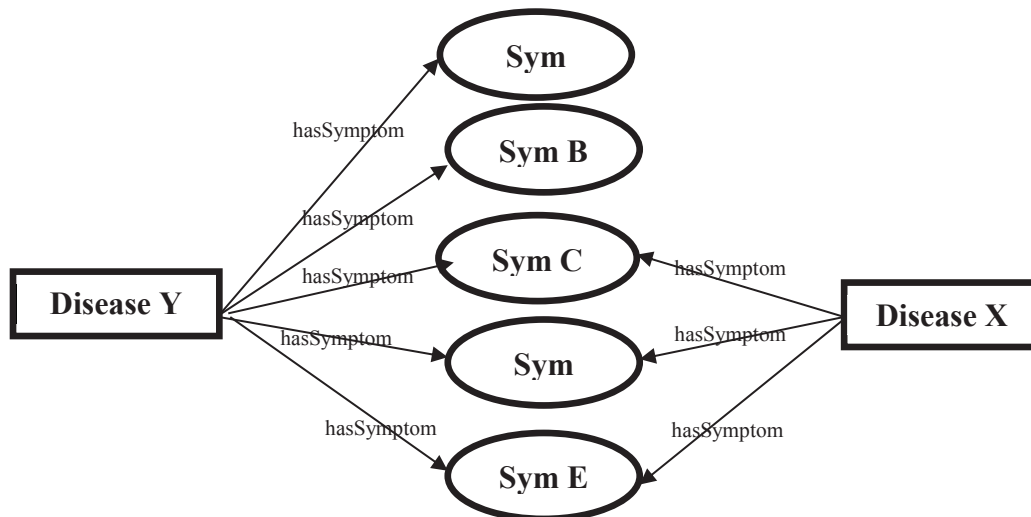


Figure 8: Nouvelle définition de la classe Disease Y.

Cette approche présente des avantages car (Rodríguez-González & Alor-Hernández, 2013) : (a) elle permet d’inférer directement la « Disease Y » en mettant ses signes (Sym A et Sym B) plus les signes de la « Disease X » (Sym C, Sym D et Sym E), (b) elle facilite la génération des règles. Il est possible de générer les règles JenaRules pour chaque classe directement sans faire de boucles récursives car toutes les définitions sont de niveaux 1 et, (c) les temps de traitement sont mieux qu’avec l’approche SeDeLo. En effet, la durée moyenne des inférences avec cette approche est d’environ 5 secondes (3 secondes après une

réinitialisation du moteur d’inférences) dans le pire des cas pour une ontologie ayant 24 maladies, 197 signes et 27 diagnostiques.

Dans une certaine mesure, la gestion des diagnostics multi-niveaux pose encore des problèmes car elle est conditionnée par la modification des ontologies. Ces modifications sont caractérisées par l’ajout de nouvelles restrictions ou par la modification des restrictions existantes (Rodríguez-González & Alor-Hernández, 2013). Ces modifications rendent ensuite les tâches de raisonnement très délicates à gérer. D’autre part, les méthodes n’ont été testées et évaluées que sur des petites ontologies (en moyenne 15 diagnostics/maladies).

### 3.3.4.2 ODDIN et les probabilités

L’approche décrite (García-crespo et al., 2010) pour ODDIN concerne une ontologie contenant les connaissances (les signes et les symptômes) sur les maladies et un fichier externe contenant les couples (maladie, observation) probabilisés (CPT) et les probabilités *a priori* (PP). Le moteur d’inférences à chaque cycle de raisonnement fait la correspondance entre les observations, le diagnostic potentiel et le fichier externe pour extraire les probabilités (CPT et PP) afin de dériver les probabilités des maladies. Les observations ici concernent l’âge et le sexe du patient, les signes cliniques, les résultats d’examen, les antécédents familiaux, etc.

Les probabilités *a posteriori*, c'est-à-dire les probabilités après la prise en compte des observations sont dérivées via la méthode de Bayes (Smets, 2005).

Les limites de cette approche sont les suivantes : (a) l’utilisation d’un fichier externe pour la prise en compte des probabilités impose l’exploitation de deux modèles distincts (l’ontologie et le fichier des probabilités). Cela entraîne des difficultés pour la maintenance de l’ontologie et des temps d’inférence plus longs. En effet, le moteur d’inférences doit faire en permanence une mise en correspondance entre les deux modèles pour extraire les CPT et PP. (b) Les systèmes probabilistes ne sont pas exportables (Seroussi et al., 2013), plus précisément les PP ne sont pas exportables. Ces systèmes doivent tenir compte des bases de données locales qui permettent de réviser (affiner) les probabilités initiales fournies par les experts du domaine.

### 3.3.4.3 Ontologie et connaissances imprécises et vagues

Les connaissances imprécises et vagues, encore appelées connaissances floues ou connaissances mal définies (Bouchon-Meunier, 2007) sont fréquentes en médecine. On pourra citer par exemple l’hypercalcémie, l’anémie, l’insuffisance rénale, etc. Dans la littérature deux approches majeures (Fernando Bobillo, 2008; Peelen et al., 2007) sont proposées pour représenter et prendre en compte ce type de connaissances lors des processus de raisonnement.

La première approche proposée par Peelen (Peelen et al., 2007) consiste à représenter ces connaissances en se référant à leurs définitions précises disponibles dans la littérature. Par exemple, l’hypercalcémie est une quantité du calcium sérique supérieure à 2,6 mmol/L (Duron, 1999), l’anémie est une quantité d’hémoglobine inférieure à 10 g/dL (de Besoist, McLean, Egli, & Cogswell, 2008). D’autres signes ou paramètres peuvent être ajoutés à ces



définitions. A ce moment-là, les extensions « *SHOIN(D)*, *SROIQ(D)*, etc. » des logiques de description classiques peuvent être utilisées.

Lorsque ces définitions ne sont pas disponibles, on utilise des approximations, par exemple « faible quantité d’Immunoglobuline de type A ». Ces approximations sont gérées par la seconde approche qui utilise la logique floue (Bouchon-Meunier, 2007), plus précisément une extension<sup>17</sup> des logiques de description implémentant la logique floue « *Fuzzy DL* » (Straccia, 1998). Fernando Bobillo (Fernando Bobillo, 2008) propose un outil « *DeLorean* » (Bobillo, Delgado, & Gómez-Romero, 2008; Bobillo, Delgado, & Gómez-Romero, 2012) mettant en œuvre la *Fuzzy DL*. *DeLorean* implémente une approche consistant à transformer une ontologie contenant des concepts flous en une ontologie classique (avec des concepts non-flous). Cette ontologie résultante peut être ensuite exploitée avec les moteurs d’inférences classiques du Web sémantique.

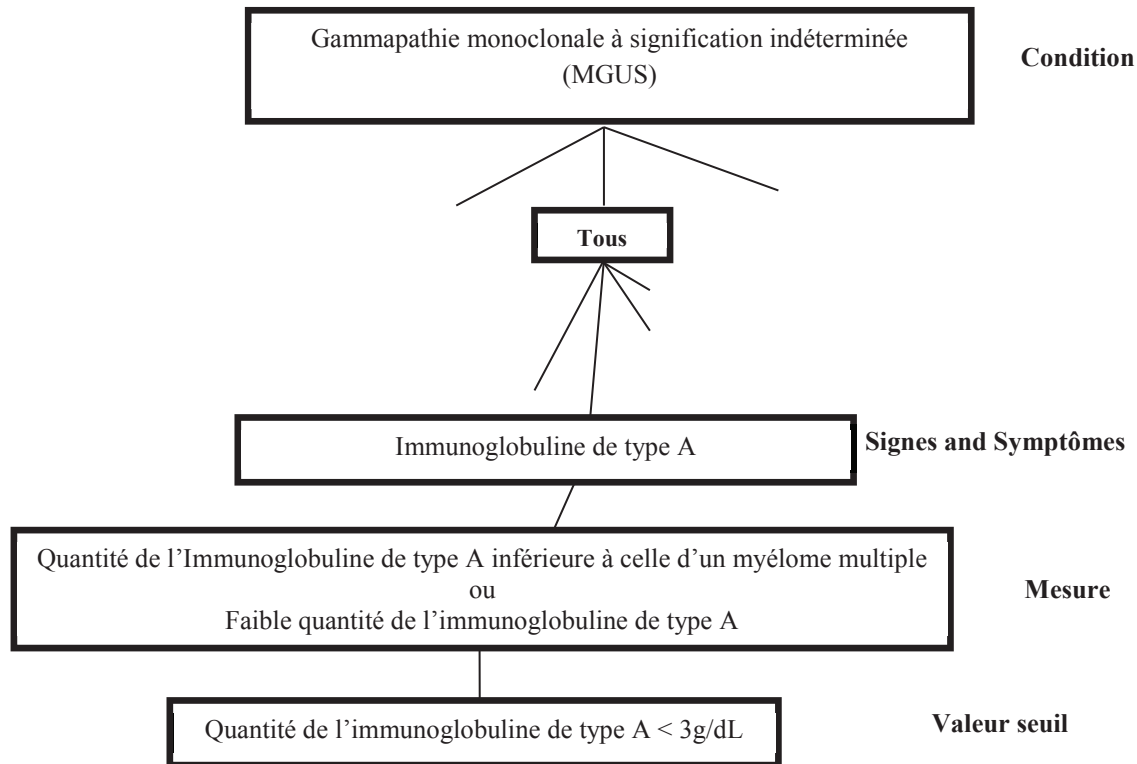
Toutefois, nous nous focalisons uniquement sur la première approche (Peelen et al., 2007) dans ce mémoire, pour plus de détails sur la seconde approche lire (Fernando Bobillo, 2008).

Peelen (Peelen et al., 2007) propose 4 niveaux d’opérationnalisation (Tableau 2)(Figure 9) : (a) le premier niveau très abstrait concerne la « condition » à observer pour qu’un patient soit atteint d’une pathologie donnée, (b) le deuxième niveau se focalise sur l’individualisation des « signes et symptômes », (c) le troisième niveau concerne l’évaluation ou la « mesure » des signes et symptômes et, (d) le dernier niveau, plus concret concerne les termes utilisés pour définir les valeurs seuils pour ces mesures.

**Tableau 2: Les 4 niveaux d’opérationnalisation des maladies** (Peelen et al., 2007).

| Niveau              | Description  | Exemple   |
|---------------------|--|---|
| Condition           | L’ensemble des signes et symptômes pour lesquels l’observation de tous ou d’une partie de ces critères implique l’observation ou non de la pathologie. | MGUS  |
| Signes et symptômes | Un critère que peut ressentir et exprimer le patient ou alors qui peut être observé par le médecin.  | Immunoglobuline de type A   |
| Mesure              | Résultat d’une évaluation faite par le médecin   | Quantité de l’Immunoglobuline de type A inférieure à celle d’un myélome multiple<br>ou<br>Faible quantité d’Immunoglobuline de type A |
| Valeur Seuil        | Ce seuil détermine si le résultat de l’évaluation indique que le signe ou le symptôme est présent ou absent  | Quantité de l’immunoglobuline de type A < 3g/dL   |

<sup>17</sup> Dans ce contexte, nous parlons de nouveaux constructeurs ajoutés au langage de base pour couvrir des besoins spécifiques. Cette notion est utilisée dans d’autres sous-sections.



**Figure 9** : Une partie de la description d'une MGUS pour mettre en exergue les niveaux d'opérationnalisation d'une ontologie (Peelen et al., 2007).

Cette approche permet de mieux qualifier les concepts mal définis lorsque les définitions précises sont disponibles. Les valeurs seuils peuvent être exprimées quantitativement ou qualitativement. Par exemple, la quantité de l'immunoglobuline de type A < 3g/dL peut être exprimée en utilisant une valeur numérique. Et pour l'exprimer qualitativement, Peelen (Peelen et al., 2007) propose de créer une hiérarchie de concepts comme ci-après :

- Quantité de l'immunoglobuline de type A
  - Quantité de l'immunoglobuline de type A < 3g/dL
  - Quantité de l'immunoglobuline de type A  $\geq$  3g/dL

Cette approche (Peelen et al., 2007) présente l'intérêt d'être facile de mise en œuvre et surtout d'être exploitable via les versions standards des langages du web sémantique.

### 3.3.4.4 Ontologie et probabilité

Notre objectif n'est pas de traiter de la nature des probabilités, à savoir subjectives ou objectives, mais plutôt de se concentrer sur les capacités des outils du web sémantique à les représenter dans un modèle ontologique. C'est-à-dire comment représenter ces valeurs réelles dans une ontologie afin de pouvoir dire qu'une maladie M a pour signe S avec une probabilité p. Tous les travaux actuels se focalisent sur les ontologies et les réseaux bayésiens (RB), à savoir, comment produire un RB à partir d'une ontologie ? A ce sujet, les travaux de Fenz (Fenz, 2012) présentent adéquatement un état de l'art dans le domaine.

Conventionnellement les ontologies contiennent des concepts et des relations organisés de façon déterministe (les relations entre concepts sont exprimées sans les probabilités). De façon générale, les relations de type « est\_un » (is\_a en anglais) sont déterministes, par exemple, la lésion lytique est une lésion osseuse et le myélome multiple est une maladie plasmocytaire. En revanche, la prévalence d’une MGUS pour une population donnée est de nature probabiliste. Il en est de même pour les relations entre les maladies et les signes ou les symptômes. Par exemple, la probabilité  $p$  de la survenue d’un signe  $S$  pour une maladie  $M$ .

Dans la littérature, plusieurs approches ont été proposées, elles reposent sur : (a) la construction de réseau bayésien (RB) à partir d’une ontologie, (b) l’extension des langages de représentation des connaissances, et (c) la création de nouvelles relations pour exprimer les connaissances probabilistes. L’utilité des RB à ce niveau se justifie pour plusieurs raisons. Les RB permettent de quantifier les croyances via des probabilités *a priori* et des probabilités conditionnelles fournies par des experts du domaine ou calculer à partir d’une base de cas existante (Kayaalp, 2005). Un RB est un graphe causal auquel on a associé une représentation probabiliste sous-jacente (Naïm, Willemin, Leray, Pourret, & Becker, 2007). Plus précisément, Richard Neapolitan (Neapolitan, 2003) définit un RB comme une large structure graphique permettant de représenter les relations probabilistes entre des variables et de faire des inférences. Un RB permet de rendre quantitatif les raisonnements sur les causalités que l’on peut faire à l’intérieur du graphe. Les inférences dans ce type de représentation consiste à calculer les probabilités conditionnelles d’événements reliés les uns aux autres par des relations de cause à effet. Il tient compte simultanément des connaissances *a priori* et des données entrées au fur et à mesure de son utilisation. Les RB fournissent un moyen de représentation de l’incertitude, c’est-à-dire un moyen de combiner les connaissances *a priori* avec des observations et une théorie d’apprentissage statistique permettant potentiellement d’affiner l’ontologie. A ces raisons, s’ajoute le fait que les RB ont une structure similaire à la structure des graphes RDF d’une ontologie en OWL car, ils sont tous des graphes orientés et ils implémentent des correspondances entre les nœuds et les arcs de leurs graphes (Ding, Peng, & Pan, 2006). Les classes et les instances d’une ontologie OWL deviennent des nœuds dans un RB et les relations d’une ontologie OWL deviennent des arcs probabilisés dans un RB.

Nous avons recensé dans la littérature trois travaux : (a) la construction des RB à partir d’une ontologie et d’une table de probabilités conditionnelles (Fenz, 2012), (b) l’extension du langage de représentation des connaissances OWL (McGuinness & Harmelen, 2004) en OWL probabiliste (PROWL) (Cesar, Costa, Laskey, & Laskey, 2005; Kathryn Blackmond Laskey & Haberlin, 2011), (c) l’ajout de nouvelles relations (Bertaud-Gounot, 2008) munies d’une sorte de facteur de certitude (tiré du système MYCIN (Buchanan & Shortliffe, 1984)).

L’approche proposée par Fenz (Fenz, 2012) consiste à construire un RB à partir d’une ontologie existante et une table de probabilités conditionnelles. Elle a fait l’objet de la mise en

place d’un *plug-in*<sup>18</sup> pour PROTÉGÉ 4.1. De manière général, un RB de  $n$  variables est un graphe orienté acyclique (DAG) de  $n$  nœuds reliés par des arcs ou flèches orientées. Un nœud  $X_i$  dans le DAG correspond aux variables, et les arcs entre deux nœuds représentent des liens causaux ou des relations d’influences d’un nœud par rapport à un autre. L’incertitude des relations causales est représentée localement par une table de probabilités conditionnelles (CPT)  $P(X_i|\Pi_i)$  associées à chaque nœud  $X_i$ , avec  $\Pi_i$  l’ensemble des parents de  $X_i$ . La distribution des probabilités de  $X = (X_1, \dots, X_n)$  peut s’exprimer sous forme de produit de probabilités conditionnelles dans le réseau (Naïm et al., 2007) :

$$P(X = x) = \prod_{i=1}^n P(X_i|\Pi_i)$$

Cette approche s’appuie sur l’utilisation d’une ontologie existante, une CPT et le logiciel Norsys Netica Bayesian Network<sup>19</sup> (NNBN) pour construire un RB. NNBN élabore un RB à partir des classes, des relations et des instances contenues dans l’ontologie. Cette construction s’appuie uniquement sur les classes, les relations et les instances contenues dans l’ontologie et se déroule en quatre étapes :

- ❖ La sélection des classes, des instances et des relations pertinentes. Cette phase est importante car elle permet à l’expert du domaine de se rassurer qu’il n’y aura pas d’informations redondantes dans le RB. Par exemple, Si A est influencé par B et B est influencé par C, les relations de B à A et de C à B seront les seules à être extraites. La relation de C à A ne sera pas extraite.
- ❖ La création de la structure du RB : le RB est automatiquement construit grâce au logiciel NNBN à partir des classes, des relations et des instances extraites précédemment.
- ❖ La construction de la CPT.
- ❖ L’intégration dans le RB des connaissances factuelles existant dans l’ontologie.

Cette approche présente deux avantages : (a) elle s’appuie sur une ontologie existante, et permet par conséquent de conserver une partie de la sémantique du domaine et (b) elle ne requiert aucune extension du langage OWL.

Cependant, son unique limite tient du fait qu’elle ne permette de faire que du raisonnement probabiliste. C’est une limite importante (pour des raisons citées précédemment) (Seroussi et al., 2013) (cf. 1.1, 2.3 et 2.3.3), si on veut utiliser le modèle qui découle de ces transformations (Fenz, 2012) pour faire du diagnostic médical (cf. 2.3).

Paulo Cesar (Cesar et al., 2005; Kathryn Blackmond Laskey & Haberlin, 2011) procèdent par extension du langage d’ontologie OWL (McGuinness & Harmelen, 2004), cette extension est PROWL (OWL probabiliste). PROWL fournit des constructeurs spécifiques pour modéliser les connaissances à caractères incertains (de nature probabiliste) dans les

<sup>18</sup> Plugin pour la construction d’un réseau Bayésien à partir d’une ontologie en OWL.  
<http://stefan.fenz.at/ontology-based-generation-of-bayesian-networks/>

<sup>19</sup> <http://www.norsys.com>

ontologies et un moteur d’inférences probabiliste. La logique de base de PROWL est celle de la Multi-Entity Bayesian Network (MEBN) (Kathryn B Laskey & Costa, 2005). Le MEBN est un système formel de représentation des connaissances intégrant la logique de premier ordre (LPO) avec la théorie bayésienne des probabilités (BPT). La représentation du MEBN s’appuie sur les entités, les attributs et les relations de l’ontologie. Les connaissances probabilistes sont exprimées sous forme de collection de MEBN fragments (MFragments). Un MFragment est une structure qui représente des connaissances probabilistes (probabilités conditionnelles) pour une collection d’hypothèses et les informations associées (par exemple les maladies et les signes). PROWL permet de capturer les connaissances d’un domaine sous forme d’un ensemble de MFragment et utilise les informations probabilistes qui y sont enregistrées pour produire des réponses aux requêtes (Kathryn Blackmond Laskey & Haberland, 2011). Selon Fenz (Fenz, 2012) le seul inconvénient de ce nouveau langage est qu’il n’est pas encore recommandé par le W3C.

Bertaud-Gounot (Bertaud-Gounot, 2008) procède par la création de nouvelles relations pour exprimer les connaissances probabilistes dans une ontologie. Bertaud-Gounot s’appuie sur les relations « *diagnoses* », « *justifies* » et crée des sous relations correspondantes, par exemple « a pour diagnostic probable » (« *probably diagnoses* ») à laquelle est rattachée un indice de certitude de +0,6. Cette technique trouve son origine dans les travaux de Bruce Buchanan et Edward Shortliffe (Buchanan & Shortliffe, 1984) sur le système MYCIN. En effet, dans MYCIN les règles et les faits ont des facteurs de certitude qui déterminent la croyance que les décideurs ont sur une règle ou un fait. Ensuite, des règles de calcul fondées sur la logique floue (Zadeh, 1965; Zalila et al., 2008) sont intégrées dans le moteur d’inférences pour agréger ces facteurs de certitude lors des inférences.

### 3.3.4.5 Ontologie et signes négatifs ou signe absent « *negative finding* »

L’expression « signe négatif » concerne tous les signes qui sont exprimés avec des négations, par exemple, « pas de douleur lombaire, absence de fièvre, pas de métastase ». C’est aussi, dans certaines situations un signe dont c’est l’absence qui a une signification, par exemple l’absence de CRAB (hypercalcémie, insuffisance rénale, anémie, lésion osseuse) est caractéristique de la gammapathie monoclonale à signification indéterminée (Rajkumar, 2011).

Des études ont permis de mettre en évidence la fréquence et l’importance des signes négatifs en médecine (Elkin et al., 2005; Pradeep & Prakash, 2001). Par exemple, une étude dans le CHU de l’Université John Hopkins a révélé que sur 14791 concepts cliniques retrouvés dans 41 dossiers médicaux<sup>20</sup>, 1823 concepts (soit environ 12,3%) ont été identifiés comme négatifs par un expert (Elkin et al., 2005).

L’importance des signes réside sur le fait qu’un dossier médical non renseigné sur des signes négatifs pertinents peut avoir des conséquences médico-légales. Ainsi, en 1998, un conseil d’experts indépendants du système national de santé britannique « *National Health System* » (NHS) appelé pour juger un cas spécifique a conclu que la tenue du dossier médical était en dessous du niveau de bonne pratique car des signes négatifs, tels que l’absence d’odeur de cétones dans l’haleine d’une patiente diabétique n’avaient pas été mentionnés (Health Service Ombudsman for England, 1998). Aux États-Unis, le règlement des systèmes

---

<sup>20</sup> Ces dossiers ont été extraits d’une base de données dans le CHU de l’Université John Hopkins

Medicare et Medicaid stipule que le médecin doit indiquer les signes négatifs anormaux et pertinents relevés lors de l’examen d’une partie du corps symptomatique (Ceusters et al., 2006, 2007).

Pour représenter les signes négatifs Bertaud-Gounot (Bertaud-Gounot, 2008) a procédé par la création de nouveaux concepts, c’est-à-dire un concept dit positif et un autre, dit négatif. Par exemple, pour représenter le fait qu’il « n’y a pas de métastase au poumon », il faut créer un concept « pas de métastase », cela donne « Entité : **poumon** » « Relation : **a** » « Entité : **pas de métastase** ».

Werner Ceusters (Ceusters et al., 2006, 2007) propose en revanche de créer une nouvelle relation « lacks, (« **n’a pas** ») » qui permet d’avoir « Entité : **poumon** » « Relation : **lacks** » « Entité : **métastase** ». Cette approche est similaire à celle utilisée par le « *National Cancer Institute Thesaurus* » (NCIT) (Golbeck et al., 2003) pour les relations de type « Excludes » qui traduisent le fait qu’une maladie exclut un signe.

### 3.3.4.6 Ontologie et raisonnement abductif

Pour rappel, le diagnostic médical est abductif (Walton, 2004). Le raisonnement abductif permet de produire la meilleure hypothèse explicative d’un fait. Dans notre cas, il consiste à produire l’hypothèse de maladie la plus significative pouvant justifier la survenue d’un signe ou d’un symptôme. En effet, un moteur d’inférences mettant en œuvre un raisonnement abductif exploite une base de connaissances contenant les triplets (maladie, signe ou symptôme, relation) et, à partir des observations (signes et symptômes) effectuées, il extrait les maladies ou hypothèses de maladies correspondantes en fonction des situations matérialisées par les relations (est une cause, est une manifestation, est un site anatomique, exclut, implique etc.). Cory Henson (Henson et al., 2011) propose une approche permettant de mettre en œuvre le raisonnement abductif dans une ontologie. Cette approche consiste à transformer l’ontologie initiale en une sorte de réseau causal.

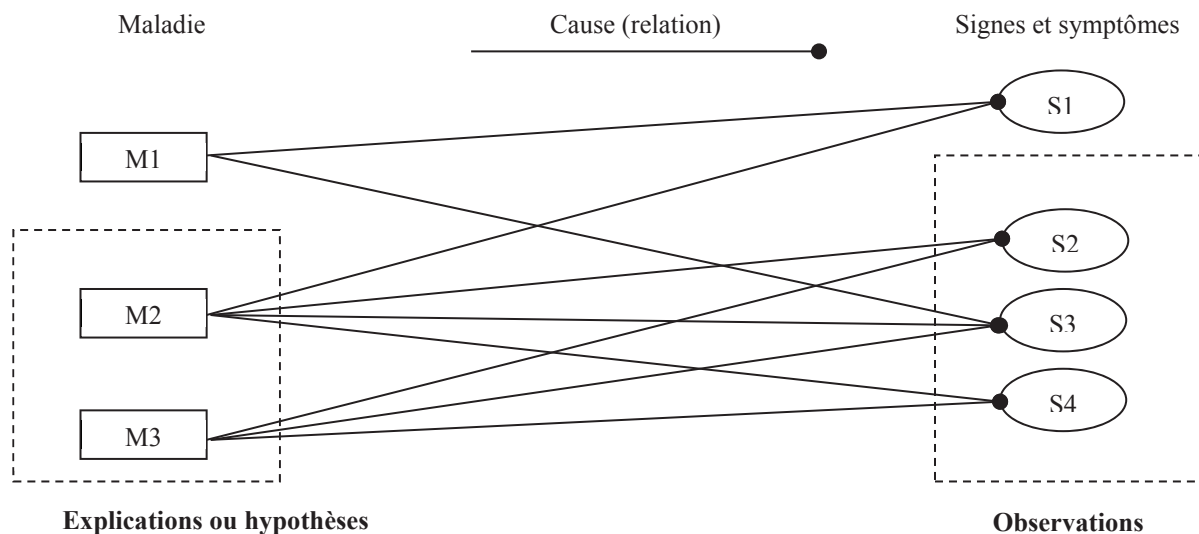


Figure 10 : Exemple de base de connaissances modélisée selon l’approche PCT (Henson et al., 2011).



En effet, selon Cory Henson (Henson et al., 2011), le langage OWL ne fournit pas de constructeurs natifs permettant de prendre en charge le raisonnement abductif. Il propose pour cela une approche mettant en œuvre le « Parsimonious Covering Theory » (PCT) (Reggia & Peng, 1987). Le PCT est un modèle formel pour le diagnostic médical dans lequel les connaissances sur les maladies sont représentées sous forme d’un réseau causal. Le but de ce modèle est de permettre au moteur d’inférences d’extraire l’hypothèse de maladie la plus plausible à partir des signes et symptômes observés. L’hypothèse la plus plausible dans ce contexte est celle ayant le moins d’observations (Sensoy, Fokoue, Srivatsa, & Pan, 2011).

Cette approche (Henson et al., 2011) présente l’avantage de ne requérir aucune extension particulière du langage OWL. Elle permet de prendre en charge lors des inférences des informations incomplètes (une partie des signes ou des symptômes décrivant une maladie). Cependant, sa limite repose sur le fait que les poids des signes pour les maladies sont les mêmes car le critère discriminant entre les hypothèses de maladies produites est le nombre d’observations les impliquant. Une approche prenant en compte les probabilités dans un moteur d’inférences abductif a été proposée par Murat Sensoy (Sensoy et al., 2011). Cette proposition consiste à s’appuyer sur le contenu de la ABox et à calculer les fréquences des observations pour chaque hypothèse de maladie produite. Ces fréquences permettent ensuite de produire les probabilités des hypothèses de maladie.

## 3.4 Redéfinition des objectifs

### 3.4.1 Problématique

Notre but principal est la modélisation des signes dans les ontologies biomédicales pour l’aide au diagnostic médical. Ce but s’appuie sur le fait que les ontologies en plus de permettre de faire de l’aide à la décision, facilitent l’intégration des données entre systèmes et l’interopérabilité entre systèmes. De ce fait, les ontologies permettent de palier une des insuffisances des précédents systèmes d’aide à la décision médicale (cf. 1.1), l’intégration au système d’information hospitalier (SIH).

En effet, l’intégration au SIH permet d’obtenir automatiquement les pré-requis (signes, symptômes, antécédents, etc.) du patient utiles pour le raisonnement du médecin. Mais avant une intégration au SIH, un système doit avoir une base de connaissances en adéquation avec les connaissances du domaine d’une part, et d’autre part, le système doit s’appuyer sur un moteur d’inférences performant. C’est pour ces raisons que nous devons proposer des méthodes générales de construction des modèles de connaissances ontologiques sur les maladies et mettre en place un moteur d’inférences permettant d’apporter une aide substantielle au médecin dans son processus du diagnostic médical.

Ces modèles de connaissances doivent contenir des connaissances du domaine (connaissances déclaratives et procédurales sur les maladies) et des connaissances de contrôle (celles liées au processus du diagnostic médical) (Clancey, 1983; Newell, 1973; Turban & Liang, 2005). Quels objectifs cherchons-nous à atteindre en mettant en place un modèle unique ? Surtout, comment intégrer toutes ces connaissances (déclaratives, procédurales et de contrôle) et les faire cohabiter dans un seul et unique modèle ?

Les systèmes existants, même ceux développés avec des modèles ontologiques ne séparent pas suffisamment les connaissances du domaine et les connaissances de contrôle. En effet, nous avons vu dans les précédentes sections que les descriptions des maladies sont modifiées afin d’être utilisables à des fins de diagnostic médical, surtout pour les cas concernant les diagnostics multi-niveaux (Rodríguez-González & Alor-Hernández, 2013). Quels sont les intérêts d’une telle séparation ? Quelle méthode ou approche mettre en œuvre pour garantir cette séparation ?

Le diagnostic médical est abductif (Pottier & Planchon, 2011; Walton, 2004) et son objectif est de produire la meilleure explication permettant de justifier la survenue d’un signe ou d’un symptôme. Ce raisonnement est incrémental car une hypothèse préalablement produite peut être confirmée ou peut être remise en cause par la survenue d’un nouveau signe ou symptôme. Alors, comment mettre en œuvre un mode de raisonnement abductif sur un modèle de connaissances ontologique ?

Selon Cory Henson (Henson et al., 2011) l’hypothèse la plus plausible produite via un raisonnement abductif est celle ayant le moins d’observations. Cela suppose que les signes ont tous les mêmes poids pour toutes les maladies, ce qui dans la réalité ne l’est pas. A ce niveau l’intégration des poids (probabilités) des signes pour les maladies dans le modèle s’avère nécessaire. Est-il possible d’intégrer des probabilités dans un modèle sous forme de PCT (cf. 3.3.4.6) ? Si oui, comment intégrer ces probabilités ?

Les méthodes probabilistes (méthode de Bayes (Aronoff, 2010)) s’appuient sur des théories bien établies (Negnevitsky, 2002b) et elles ont fait l’objet de plusieurs mises en œuvre (De Dombal et al., 1972; Mendonça, 2004; Miller, 2009; Vikram & Karjodkar, 2009). Cependant, nous devons tenir compte du fait que ces modèles ne sont pas transférables (exportables) (Seroussi et al., 2013) et requièrent toujours une phase d’apprentissage (cf. 1.1). L’utilisation de la méthode de Bayes requiert la présence de certains paramètres, à savoir la prévalence des maladies, la sensibilité et la spécificité des signes pour les maladies (cf. 2.4). La production de la spécificité d’un signe pour une maladie requiert la prise en compte des vrais négatifs (VN). Cet effectif (VN) est lié aux signes absents. Comment intégrer ces signes absents ? Et, comment les manipuler ?

### 3.4.2 Objectifs

Nous voulons proposer et valider des méthodes générales de construction d’ontologies biomédicales permettant de faire du diagnostic médical et mettre en place un moteur d’inférences en utilisant les technologies du web sémantique.

#### 3.4.2.1 Modèle de connaissance

Ces ontologies doivent tenir compte de la typologie des connaissances définie par Grundstein (Grundstein et al., 1988a) :

- ❖ **Les connaissances déclaratives.** Par exemple, un MM est une maladie plasmocytaire (Rajkumar, 2011). La prévalence d’une MGUS pour les personnes de plus de 70 ans est de 5% (Kyle, Francis, & Vincent, 2011; Watanaboonyongcharoen, Nakorn, Rojnuckarin, Lawasut, & Intragumtornchai, 2012).
- ❖ **Les connaissances procédurales.** Par exemple, si la patient a une quantité d’immunoglobuline de type A < 3g/dL, Alors penser à une MGUS (Prcela et al.,



2008). On peut aussi avoir des règles permettant de calculer les rapports de vraisemblance à partir des sensibilités et spécificités des signes pour les maladies.

- ❖ **Les connaissances de contrôle.** Par exemple, pour les signes « présents » utiliser ce type de connaissances, sinon utiliser tels autres types de connaissances. On peut aussi avoir des règles utiles pour définir les catégories de signes (Signes biologiques, signes radiologiques, etc.).

A partir des modèles ontologiques élaborées nous produirons des ensembles de règles qui seront utiles pour les moteurs d'inférences dans la production des hypothèses de maladies.

### 3.4.2.2 Système d'aide à la décision à base d'ontologie (ODES)

Les ensembles de règles produits seront exploités via un système d'aide à la décision. Ce système permettra de présenter le contenu des modèles ontologiques à un praticien afin de l'aider dans ces tâches. Ce système doit fournir une interface adaptative (Frank & Szekely, 1998) en fonction du contenu de l'ontologie. La structure du contenu de l'interface est définie par certaines connaissances de contrôle. Ce système doit aussi fournir des explications (Negnevitsky, 2002c) sur son raisonnement guidées par les connaissances de contrôle.

## 3.5 Conclusion

Dans ce chapitre, nous avons présenté les typologies des connaissances via la différenciation des connaissances en elles-mêmes (déclarative, procédurale et contrôle). Nous avons ensuite présenté les langages de représentation des connaissances et les travaux existants dans la littérature traitant des ontologies. Il ressort de cela que le modèle ontologique permet potentiellement de mettre en œuvre un modèle de connaissances sur les maladies prenant en compte la complexité du domaine médical. En effet, les ontologies donnent la possibilité :

- ❖ d'avoir un modèle de connaissances sur les maladies via la définition des relations entre les maladies et les signes, comme fait par García-crespo dans ODDIN (García-crespo et al., 2010), Charles le MOING (Le Moing, 2009) et Bertaud-Gounot (Bertaud-Gounot, 2008).
- ❖ d'intégrer des règles (Jovic, Prcela, & Gamberger, 2007; Prcela et al., 2008).
- ❖ de pouvoir formaliser le raisonnement médical via l'ajout de nouvelles relations (Bertaud-Gounot, 2008), ou par la création d'un réseau causal suivant le modèle PCT (Henson et al., 2011).
- ❖ de pouvoir représenter les probabilités (Fenz, 2012).
- ❖ d'avoir un modèle réutilisable (Gruber, 1995; Neches et al., 1991; Studer et al., 1998).
- ❖ de pouvoir représenter les signes négatifs ou absents (Bertaud-Gounot, 2008; Ceusters et al., 2006, 2007).

Et enfin, nous avons présenté la problématique et les objectifs détaillées de ce travail. Tout au long des chapitres qui vont suivre nous proposerons des réponses aux questions présentées ci-haut.

Pour la suite, les noms et les contenus des figures, des équations, des classes, des règles et des relations respectent les critères de nommages du National Cancer Institute Thesaurus (NCIT) (Ceusters, Barry, & Goldberg, 2005; Fragoso, de Coronado, Haber, Hartel, & Wright, 2004), comme sur les Figure 11 et Figure 12.

## Chapitre 4

### Matériels et méthodes

#### 4.1 Matériels

Nous présentons dans cette section le matériel que nous avons utilisé pour mettre en place les méthodes de modélisation des signes dans les ontologies biomédicales et les mécanismes de raisonnement associés pour l'aide au diagnostic.

##### 4.1.1 Le corpus

Trois domaines d'application ont été sélectionnés, à savoir : les maladies plasmocytaires, les urgences odontologiques et les lésions traumatiques du genou. Les articles scientifiques sélectionnés sont tous récents (moins de 10 ans) et traitent de l'état des connaissances dans ces domaines. Ils contiennent les définitions opérationnelles des pathologies, les approches thérapeutiques, etc.

Ces ressources utilisées pour la mise en place des méthodes incluent les taxonomies existantes telles que l'ontologie de la « *National Cancer Institute Thesaurus* » (NCIT) (Ceusters et al., 2005; Fragoso et al., 2004), la recommandation de la *Haute Autorité de la Santé* (HAS) intitulée « Recommandations professionnelles : Prise en charge thérapeutique des lésions méniscales et des lésions isolées du ligament croisé antérieur du genou chez l'adulte » (Haute Autorité de Santé publique, 2008) et les définitions opérationnelles :

- ❖ des maladies plasmocytaires (Decaux et al., 2006; Kyle et al., 2011; Rajkumar, 2011) et,
- ❖ des pathologies en urgences odontologiques (Bertaud-Gounot et al., 2010; Kalendarian et al., 2011; Le Moing, 2009; Seon Gyu, Hong-Gee, & Myeng-Ki, 2007; Seon Gyu & Hong-Gee, 2006).

A cela s'ajoute des dossiers patients qui seront utilisés pour tester et évaluer les modèles mis en place.

##### 4.1.1.1 L'ontologie du NCIT

Le « *NCI Thesaurus* » est une ontologie (Figure 11) de domaine développée et publiée par l'« *Enterprise Vocabulary Services* » (EVS) un consortium entre le National Cancer Institute (NCI) office et le « *NCI Center for Bioinformatics* » (NCICB). C'est une terminologie de référence qui couvre tout le domaine de la cancérologie. Elle a été conçue pour faciliter le partage des connaissances et la recherche dans le domaine de la cancérologie. Elle contient approximativement 110 000 termes, dont 36 000 concepts, subdivisés en 20 sous domaines, incluant les maladies, les médicaments, l'anatomie, les gènes, les produits de gènes, les processus biologiques, etc. Chaque concept représente une unité de sens et contient des annotations, des synonymes et les labels préférés. Les concepts sont représentés en logique de description (LD) et définis par des relations avec d'autres concepts. Il existe approximativement 90 types de relations dans cette ontologie. L'intérêt de cette ontologie repose sur le fait qu'elle contient plusieurs types de relations qui représentent la sémantique

du domaine médical (Ceusters et al., 2005; Fragoso et al., 2004). Des types de relations (Figure 12) tels que :

- ❖ « Is a », relation taxonomique,
- ❖ « Part of », relation de méronomie,
- ❖ « Has », qui exprime le fait qu'une maladie M donnée a pour signe S, site anatomique S1 ou cellule anormale S2, etc. Cette relation stipule aussi que S est constant pour M, et que par conséquent si la présence de S confirme la présence de M, alors l'absence de S infirmera la présence de M.
- ❖ « Excludes », qui exprime le fait qu'une maladie M exclut le signe S, site anatomique S1 ou cellule anormale S2. Cette relation s'apparente à la relation « lacks » introduite par Werner Ceusters (Ceusters et al., 2007) dans l'« *Open Biological and Biomedical Ontology* » (OBO). Cette relation stipule aussi que l'absence de S est constant pour M, et que par conséquent on peut dire que si la présence de S infirme la présence de M, alors l'absence de S confirmera la présence de M.
- ❖ « May\_Have », qui exprime le fait qu'une maladie M peut avoir pour signe S, site anatomique S1 ou cellule anormale S2. Cette relation stipule aussi que S est inconstant pour M, et que par conséquent si la présence de S confirme la présence de M, alors l'absence de S n'infirmera pas la présence de M, car l'inconstance de S fait qu'il peut être présent pour M ou non.



Figure 11: Une vue d'une partie de l'ontologie du NCIT dans l'éditeur Protégé 3.4.8.

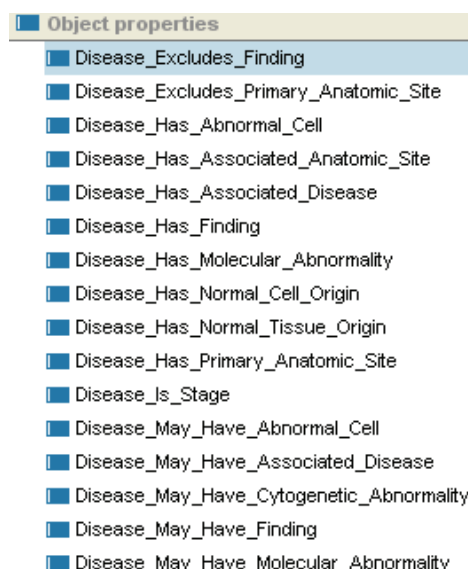


Figure 12: Une vue des relations de l'ontologie du NCIT dans Protégé 3.4.8.

Des travaux ont été menés par Natalya Noy et ses collègues (Noy et al., 2008) sur le traitement des relations de type « May\_Have », car elles expriment le fait qu'une maladie peut avoir un signe ou non, pour illustrer cela, ils prennent le cas suivant :

*Astroblastoma* :

**Disease\_May\_Have\_Finding** *some Necrotic\_Change*

En effet, cette relation stipule que certaines instances de la maladie « Astroblastoma » ont le signe «Necrotic\_Change» et d'autres ne l'ont pas ou même, il est possible que certaines l'excluent. La solution proposée par Natalya Noy et ses collègues (Noy et al., 2008) consiste à faire en sorte que les relations de type « May\_Have » subsument les relations de types « Has » (Figure 13)

Figure 13. Cette solution permet d'exprimer les cas où un signe serait optionnel au niveau d'une superclasse, mais obligatoire au niveau des sous-classes.

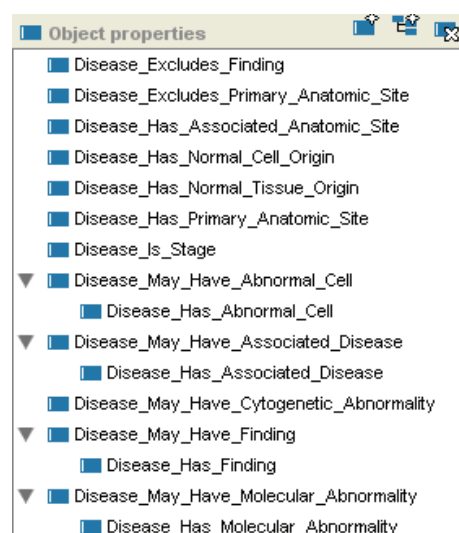


Figure 13: Une vue des relations de l'ontologie du NCIT dans Protégé 3.4.8 après la modification sur la hiérarchie des relations apportée par Natalya Noy et ses collègues (Noy et al., 2008).

Toutes les relations (Tableau 3) d'une ontologie sont caractérisées par un ensemble de départ (domaine) et un ensemble d'arrivée (range ou co-domaine). Chaque relation lie une instance I1 d'un domaine D à une instance I2 d'un co-domaine R.

Tableau 3: Quelques relations de l'ontologie du NCIT

| Nom                      | Domaine                | Co-domaine |
|--------------------------|------------------------|------------|
| Disease Has Finding      | Diseases and Disorders | Finding    |
| Disease Excludes Finding | Diseases and Disorders | Finding    |
| Disease May Have Finding | Diseases and Disorders | Finding    |

#### 4.1.1.2 Le domaine d'application N° 1 : les maladies plasmocytaires

Le myélome multiple est une lymphopathie B due à une expansion clonale de plasmocytes matures. Au plan tissulaire, on retrouve trois types de lésions, il existe : (a) une augmentation de la cellularité de la moelle liée à l'infiltration plasmocytaire mais aussi à la réactivation de la moelle rouge ; (b) une néo-angiogenèse qui se développe, grâce aux cytokines pro-angiogéniques ; (c) une stimulation des ostéoclastes par les cellules myélomateuses responsables de lésions ostéolytiques focales et d'une ostéopénie diffuse (Baur, Bartl, Pellengahr, Baltin, & Reiser, 2004). L'évolution des lésions tissulaires se fait de manière séquentielle parallèlement à l'évolution clinique de la maladie (Sirohi & Powles, 2004).

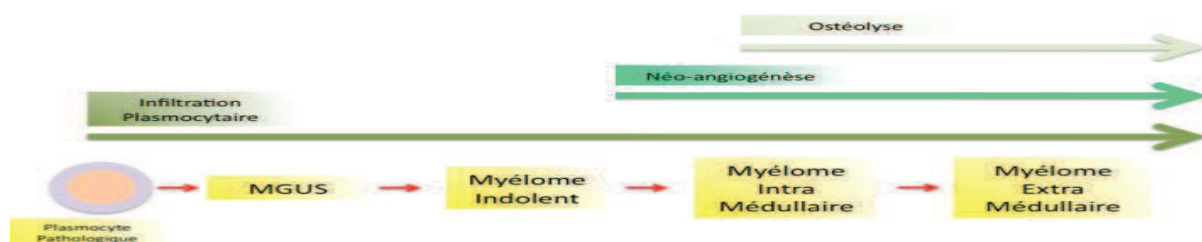


Figure 14 : L'évolution clinique du patient se fait parallèlement aux modifications histologiques (Sirohi & Powles, 2004).

Le développement des néo-vaisseaux est une étape importante car ils représentent la transition entre le myélome indolent (*smoldering myeloma* pour les anglo-saxons) et le myélome intra-médullaire symptomatique, aggravant le pronostic de la maladie (Alexandrakis, Passam, Dambaki, Pappa, & Stathopoulos, 2004; Jakob et al., 2006; Moehler, Ho, Goldschmidt, & Barlogie, 2003). Ce passage à la forme symptomatique justifie la prise en charge thérapeutique et mérite d'être définie. Outre cette classification en gammopathie monoclonale à signification indéterminée (MGUS), « *smoldering myeloma* » et myélome symptomatique, on classe les myélomes classiquement en stades de Durie et Salmon (Angtuaco, Fassas, Walker, Sethi, & Barlogie, 2004) : le stade 1 correspond presque au « *smoldering myeloma* » mais il existe cependant un recouvrement avec le myélome symptomatique. Les stades 2 et 3 expriment des niveaux d'aggravation successifs et sont en rapport avec la masse tumorale (Durie, 2006). Pour les myélomes symptomatiques, on utilise aussi une classification pronostique qui s'appelle ISS (Munshi, 2008). Enfin, une nouvelle classification prenant en compte les données de l'imagerie, la classification Durie et Salmon PLUS (Durie, 2006) semble avoir un intérêt pour le suivi des patients sous traitement (Carty,

Shortt, Shelly, Eustace, & O'Connell, 2010; Lin et al., 2010). Les concepts du myélome sont bien représentés dans l'ontologie du NCIT, cependant on peut regretter certains manquements et certaines incohérences (Bertaud-Gounot, Donfack Guefack, Brilletet, & Duvauferrier, 2012; Donfack Guefack, Bertaud-Gounot, Duvauferrier, et al., 2012) :

- ❖ une MGUS y est définie comme étant un état cancéreux. Or, dans la littérature une MGUS est définie comme état précancéreux, avec une quantité d'immunoglobuline de type A < 3g/dL, un taux de plasmocyte dans moelle osseuse < 10% et une absence d'atteintes osseuses, d'hypercalcémie, d'anémie et d'insuffisance rénale (CRAB) (Decaux et al., 2006; Kyle et al., 2011; Rajkumar, 2011).
- ❖ l'absence des atteintes osseuses, d'hypercalcémie, d'anémie et d'insuffisance rénale caractérisant une MGUS n'est pas représentée.
- ❖ la notion de myélome symptomatique qui n'est pas explicitée.
- ❖ la classification Durie et Salmon PLUS très récente n'est pas non plus présente. Par ailleurs, certains critères diagnostiques ne sont pas suffisamment précis : on parlera d'anémie sans préciser la valeur de l'hémoglobine. Certains de ces manques sont liés au manque d'expressivité de la version de OWL avec laquelle a été écrit le NCIT (Fragoso et al., 2004). D'autres manques sont liés à la finalité de l'ontologie qui n'est pas conçue pour représenter les manifestations des maladies mais plutôt le cadre nosologique de la maladie (processus physiopathologiques, sites anatomiques, etc.).

L'intérêt de ces articles (Angtuaco et al., 2004; Carty et al., 2010; Decaux et al., 2006; Durie, 2006; Fragoso et al., 2004; Kyle et al., 2011; Lin et al., 2010; Munshi, 2008; Rajkumar, 2011) et de la sous-ontologie du NCIT concernant les maladies plasmocytaires est qu'ils montrent clairement que les définitions opérationnelles des maladies diffèrent des définitions ontologiques (Burgun et al., 2005). Il montre aussi que pour faire du diagnostic médical, une ontologie doit être opérationnalisée (Peelen et al., 2007) en prenant en compte les données de la littérature (Rajkumar, 2011). Selon Martyn Shuttleworth l'opérationnalisation (Shuttleworth, 2013) est : « un ensemble de tâches qui consiste à rendre opérationnel ». En d'autres termes, pour ce qui nous concerne, elle consiste à appliquer des transformations pour rendre une ontologie utilisable à des fins de diagnostic médical. Les travaux de Peelen (Peelen et al., 2007)(cf. 3.3.4.3) montrent les différentes tâches à exécuter pour opérationnaliser une ontologie biomédicale.

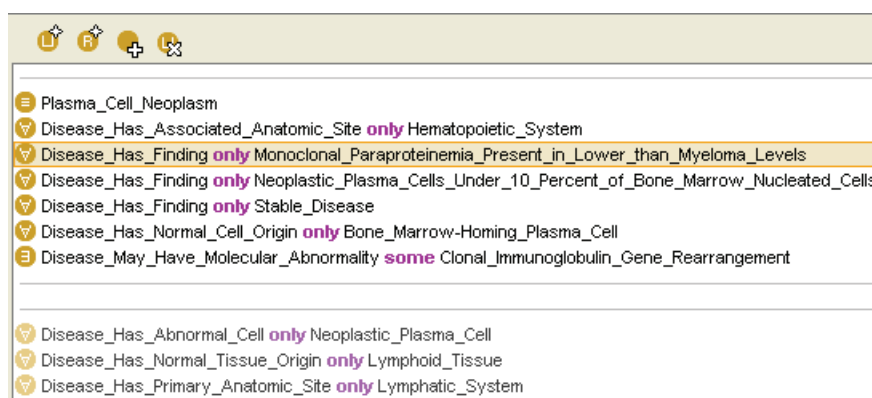


Figure 15: Définition d'une MGUS dans le NCIT (Fragoso et al., 2004). On peut constater l'absence des critères CRAB, la définition très vague de la quantité de l'immunoglobuline (ligne sélectionnée).



| Disorder  | Disease definition   |
|---|--|
| Monoclonal gammopathy of undetermined significance (MGUS) | <p>All three criteria must be met:</p> <ul style="list-style-type: none"> <li>● Serum monoclonal protein &lt;3 g/dL</li> <li>● Clonal bone marrow plasma cells &lt;10%</li> <li>● Absence of end-organ damage such as hypercalcemia, renal insufficiency, anemia, and bone lesions (CRAB) that can be attributed to the plasma cell-proliferative disorder; or in the case of IgM MGUS, no evidence of anemia, constitutional symptoms, hyperviscosity, lymphadenopathy, or hepatosplenomegaly that can be attributed to the underlying lymphoproliferative disorder.</li> </ul> |

**Figure 16: Définition d'une MGUS dans la littérature** (Kyle et al., 2011; Rajkumar, 2011).

La formalisation des critères diagnostiques est importante pour une aide au diagnostic médical à cause : (a) de la multiplicité des classifications et des critères diagnostiques, et, (b) du fait que c'est la classification (maladie) précise du patient qui va conditionner la décision thérapeutique.

#### 4.1.1.3 Le domaine d'application N° 2 : Les urgences odontologiques

L'urgence dentaire est un motif de consultation très fréquent, aussi bien dans le cadre hospitalier qu'en cabinet privé. Il n'est pas rare que la pathologie dentaire demeure longtemps asymptomatique, les symptômes n'apparaissant que lorsque la pathologie est déjà entrée en phase aigüe. Il existe une grande diversité de pathologies buccales et plus particulièrement dentaires. Il est possible de rencontrer plusieurs dizaines de pathologies de la muqueuse buccale, avec de multiples étiologies. A cause de cette symptomatologie bucco-dentaire riche et complexe, le risque de confusion des pathologies est donc important, surtout dans un contexte de stress et de surcharge de travail.

Des travaux dans le domaine odontologiques ont déjà été faits, notamment les travaux : (a) de Seon Gyu sur une ontologie des dents manquantes (Seon Gyu & Hong-Gee, 2006) et une ontologie de la position des dents (Seon Gyu et al., 2007), (b) de Charles Le Moing (Le Moing, 2009) et, (c) de Bertaud-Gounot (Bertaud-Gounot et al., 2010) sur une ontologies des urgences odontologiques.

#### 4.1.1.4 Le domaine d'application N° 3 : Les lésions méniscales et des lésions isolées du ligament croisé antérieur du genou chez l'adulte

Le genou est constitué de plusieurs éléments : les ligaments collatéraux, les ligaments croisés, le ménisque, etc. Une lésion du genou peut toucher plusieurs éléments en même temps. L'ontologie (Figure 17) contient la représentation de différents types de lésions et les signes cliniques associés (Tableau 4). Cette ontologie inclut aussi plusieurs relations (Ceusters et al., 2005; Fragoso et al., 2004) telles que :

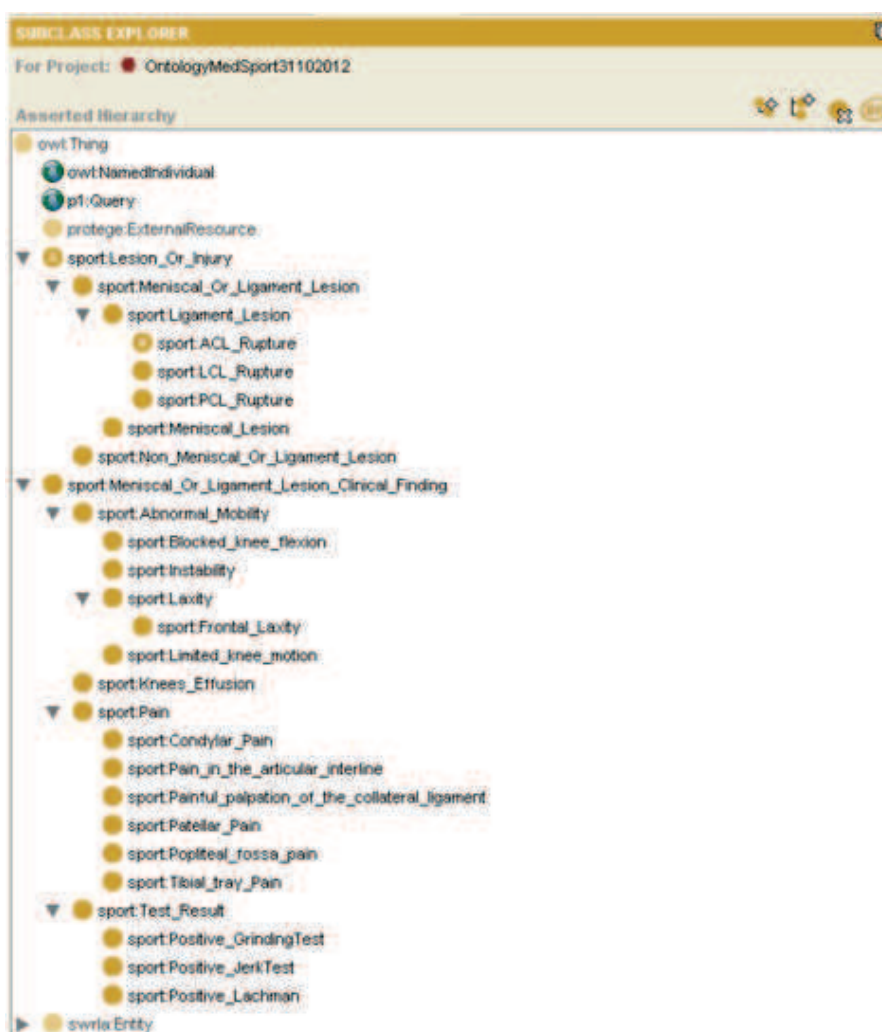
- ❖ « Has\_Finding », qui exprime le fait qu'une lésion du genou d a pour signe f.
- ❖ « Excludes\_Finding », qui exprime le fait qu'une lésion du genou d n'a jamais pour signe f.
- ❖ « May\_Have\_Finding », qui exprime le fait qu'une lésion du genou d peut avoir pour signe f.

Ces relation sont cruciales dans le processus du diagnostic médical parce qu'elles permettent l'expression des règles utilisées par les médecins ou les experts du domaine (cf. 4.2.3). Les prévalences des lésions pour une population de patients donnée et les sensibilités

et spécificités des signes pour les lésions (Tableau 4) ont été produites grâce à une étude rétrospective en milieu hospitalier.

Dans cette étude, 1000 patients ont été vus en urgence pour un traumatisme du genou et 134 d'entre eux ont été réévalués 5 jours après le traumatisme dans le service de médecine du sport. Cette étude s'est faite dans le cadre d'une évaluation des pratiques professionnelles (Haute Autorité de Santé publique, 2008).

Dans ce contexte, le Tableau 4 des probabilités associées à l'ontologie nous permettra de tester les méthodes de modélisation des connaissances probabilistes dans l'ontologie que nous allons compléter.



**Figure 17 : Ontologie des signes cliniques des lésions traumatiques du genou. Cette version de l'ontologie ne représente que les connaissances qualitatives telles que : les lésions et les signes. Des catégories de lésions, de signes et de symptômes sont aussi représentées.**

Le Tableau 4 suivant contient les types de lésions et les signes cliniques associés traités dans l'étude. Les correspondances en anglais sont aussi fournies (Wong, Khan, Jayadev, Khan, & Johnstone, 2012).



**Tableau 4: Prévalences de 4 types de lésions avec les sensibilités et spécificités (se/sp) des signes cliniques pour les types de lésions.**

| <b>Examination Sign<br/>(% Prevalence)</b>                        | <b>Lésion<br/>(Lesion)<br/>(44)</b> | <b>Lésion ligaments<br/>(Ligament lesion)<br/>(28)</b> | <b>Lésion LCA<br/>(ACL Rupture)<br/>(22)</b> | <b>Lésion IRM Ménisque<br/>(Meniscal lesion)<br/>(18)</b> |
|---|-------------------------------------|--|--|---|
| Douleur de la patella (Patellar pain)                             | 0,02/0,9                            | 0,01/0,87  | 0,01/0,90                                    | 0,04/0,96   |
| Instabilité (Instability)   | 0,07/1                              | 0,07/1   | 0,03/0,90                                    | 0,08/0,96   |
| douleur du creux poplité (Popliteal fossa pain)                   | 0,04/1                              | 0,05/1   | 0,06/1                                       | 0,04/0,96   |
| douleur du condyle (Condylar pain)                                | 0,16/0,6                            | 0,15/0,64  | 0,2/0,7                                      | 0,16/0,75   |
| Blocage (Blocking Knee flexion)                                   | 0,07/1                              | 0,05/0,92  | 0,03/0,90                                    | 0,12/1  |
| Réduction d'amplitude (Limited Knee motion)                       | 0,11/0,8                            | 0,07/0,71  | 0,06/0,77                                    | 0,12/0,85   |
| Plateau tibial (Tibial tray)                                      | 0,04/0,9                            | 0,05/0,92  | 0,03/0,90                                    | 0,08/0,96   |
| Grinding test (Positive grinding test)                            | 0,09/0,9                            | 0,07/0,85  | 0,03/0,81                                    | 0,16/0,96   |
| Interligne articulaire (Joint space)                              | 0,21/1                              | 0,18/0,57  | 0,83/0,33                                    | 0,37/0,85   |
| Jerk test (Positive Jerk Test)                                    | 0,33/0,9                            | 0,65/0,14  | 0,4 /0,86                                    | 0,29/0,71   |
| Epanchement (Knees effusion)                                      | 0,66/0,6                            | 0,71/0,64  | 0,73/0,54                                    | 0,54/0,32   |
| Laxité (Positive Lachman)   | 0,51/0,77                           | 0,62/0,84  | 0,6/0,65                                     | 0,45/0,46   |
| Ligaments latéraux (Painful palpation of the collateral ligament) | 0,38/0,4                            | 0,39/0,5   | 0,3/0,65                                     | 0,33/0,5  |

#### 4.1.1.5 Dossiers patients

Pour la validation des modèles et l'évaluation du système, des dossiers patients anonymisés ont été utilisés. (a) 63 dossiers patients fournis par le service de médecine interne de l'hôpital sud de Rennes pour l'ontologie des maladies plasmocytaires. (b) 202 dossiers patients fournis par l'unité fonctionnelle de Santé Publique et les urgences odontologiques du CHU de Rennes pour l'ontologie des urgences odontologiques. Et, (c) 20 dossiers patients fournis par le service de médecine du sport du CHU de Rennes pour l'ontologie des signes cliniques et des lésions traumatiques du genou.

#### 4.1.2 Langage et outils de représentation des connaissances

Les éditeurs d'ontologies utilisés sont Protégé 3.4.8 (ProtegeWiki, 2012) et Protégé 4.x. Le langage de représentation des connaissances dans nos ontologies est « *OWL* » dont les constructeurs sont présentés dans le tableau 5 suivant. Le langage de règles utilisé est « *SWRL* » (I. Horrocks et al., 2004).

**Tableau 5: Liste des constructeurs de la logique de description utilisée** (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2003).

| Constructeur                    | Syntaxe               | Sémantique  |
|---------------------------------|-----------------------|---|
| Nom de concept                  | $C$                   | $C'$  |
| Négation de concepts            | $\neg C$              | $\Delta' \setminus C'$  |
| Conjonction                     | $C_1 \cap C_2$        | $C_1' \cap C_2'$  |
| Disjonction                     | $C_1 \cup C_2$        | $C_1' \cup C_2'$  |
| Quantificateur universel        | $\forall R. C$        | $\{d_1 \in \Delta' \mid \forall d_2 \in \Delta'. (R'(d_1, d_2) \rightarrow d_2 \in C')\}$ |
| Quantificateur existentiel typé | $\exists R. C$        | $\{d_1 \in \Delta' \mid \exists d_2 \in \Delta'. (R'(d_1, d_2) \wedge d_2 \in C')\}$      |
| Nom de rôles                    | $R$                   | $R'$  |
| Conjonction de rôles            | $R_1 \sqcap R_2$      | $R_1' \cap R_2'$  |
| Hierarchie des rôles            | $R_1 \sqsubseteq R_2$ | $R_1' \subseteq R_2'$   |

### 4.1.3 Langages et outils de développement

Le langage « *JAVA* » est utilisé pour les développements spécifiques. L'intérêt de ce langage repose sur le fait que la majorité des outils et *plugins* exploités dans le domaine du Web sémantique sont écrits en *JAVA*.

Les API « *OWL* » (Horridge & Bechhofer, 2009), « *JESS* » (Friedman-Hill, 2008), « *SWRL* » (O'Connor, Nyulas, Shankar, Das, & Musen, 2008), « *SQWRL* » (O'Connor & Das, 2009) et « *SWRL-IQ* » (Elenius, 2012; Ford, 2012) tous écrits en *JAVA* sont utilisées pour les manipulations de nos modèles ontologiques.

La système de gestion de base de données relationnelles (SGBD) « *MYSQL* » (Oracle, 2011) est utilisé pour les tests et les évaluations de la méthode. Nous avons utilisé ce SGBD pour faciliter les tests et les évaluations multicentriques. L'éditeur de développement intégré utilisé est « *Eclipse Juno* » (Eclipse-Foundation, 2012).

#### 4.1.3.1 JESS

Le moteur de règles *JESS* permet de gérer les règles *SWRL* (Friedman-Hill, 2008). *JESS* fonctionne en chaînage avant et arrière, sauf que pour le cas du chaînage arrière les types de faits à utiliser doivent être définis. Dans ce contexte, les règles exécutées sont celles rattachées à ces types de faits déclarés. Le fonctionnement en chaînage avant est en natif dans *JESS* et le chaînage arrière est simulé.

#### 4.1.3.2 Semantic Query Enhanced Web Rule Language: SQWRL

*SQWRL* « *Semantic Query-Enhanced Web Rule Language* » est un langage de requête fondé sur le langage *SWRL* pour les ontologies faites en *OWL* (O'Connor & Das, 2009). Il fournit un ensemble de fonctions semblable à du « *Sql* » pour extraire des connaissances d'une ontologie en *OWL*. Ces requêtes peuvent être écrites via la *SWRLTab* de PROTÉGÉ.

### 4.1.3.3 Semantic Web Rule Language Interrogation Query : SWRL-IQ

SWRL-IQ « *Semantic Web Rule Language –Interrogation Query* » (Abburu, 2012; Elenius & Riehemann, 2012; Ford, 2012) est un *plugin* développé par le « *SRI International* » (Elenius, 2012) pour Protégé 3.x<sup>21</sup> qui permet de modifier, de sauvegarder et d'exécuter des requêtes en s'appuyant sur un moteur d'inférences de type XSB Prolog. Il fournit :

- ❖ un moteur d'inférences en chaînage arrière.
- ❖ un système de sauvegarde des requêtes sous format XML ou CSV.
- ❖ un journal pour la trace des exécutions.
- ❖ un système d'export des résultats sous différents formats.
- ❖ un système de gestion des relations ayant des arités supérieures à 2.

SWRL-IQ possède de nombreuses extensions, cependant il ne prend pas en charge le raisonnement sur la TBox, c'est-à-dire qu'il ne peut pas par exemple, effectuer des tâches de détection de relation de subsomption de classes (Elenius, 2012).

## 4.2 Méthodes

Nous savons que les définitions des maladies contenues dans les ontologies biomédicales classiques diffèrent des définitions contenues dans la littérature (Bertaud-Gounot, Duvauferrier, et al., 2012; Burgun et al., 2005), donc mettre en place un modèle ontologique adapté au diagnostic médical passe par un processus dit d'opérationnalisation (Peelen et al., 2007). L'opérationnalisation consiste à rendre opérationnel, rendre apte à fonctionner de manière productive (Shuttleworth, 2013). Dans notre contexte, ce processus consiste à transformer une ontologie biomédicale (ontologie contenant uniquement les connaissances sur les maladies, c'est-à-dire la nosologie) en ontologie sémiologique biomédicale (c'est-à-dire une ontologie biomédicale adaptée pour le diagnostic médical). En effet, les connaissances contenues dans l'ontologie biomédicale sont de type physiopathologique, anatomique, biologique, radiologique. Les connaissances contenues dans l'ontologie sémiologique biomédicale sont de type physiopathologique, anatomique, biologique, radiologique, manifestation (symptôme), critère diagnostique (par exemple m signes parmi n), contrôles, etc. Dans cette optique, nous traitons des connaissances (Cléret, Le Beux, & Le Duff, 2001; Fagot-Largeault, 2010) :

- ❖ *qualitatives non ordonnées*, dites nominales ou catégorielles. Ainsi, des lésions osseuses sont présentes ou absentes, une douleur du dos peut être accentuée par un changement de position ou un effort. La première est dichotomique ou bivaluée car elle ne peut prendre que deux valeurs tandis que la seconde est multivaluée.
- ❖ *qualitatives ordonnées*, dites ordinales. Il existe alors une hiérarchie dans les catégories comme, par exemple, pour évaluer la sévérité d'un myélome multiple qui, selon la classification de Durie et Salmon (Durie, 2006), est cotée de I à III.
- ❖ *quantitatives*, dites mesurables qui s'expriment par un nombre. Elles sont continues lorsqu'elles peuvent prendre toutes les valeurs comprises dans l'intervalle des variations, on peut citer la taille, le poids, la mesure de l'immunoglobuline de type A ou la concentration sérique de la créatinine. Elles sont discontinues ou discrètes

<sup>21</sup> Désigne toutes les sous-versions de Protégé 3.

lorsqu'elles ne peuvent prendre que certaines valeurs, souvent des nombres entiers comme la fréquence cardiaque par minute ou le nombre de paquet-année chez un fumeur.

Nous traitons aussi de la représentation :

- ❖ des relations d'arité  $n$  (c'est-à-dire les relations ayant des arités supérieures à 2) et les notions de réification (cf. 3.3.2.4) pour la prise en compte des connaissances probabilistes et multivaluée.
- ❖ des signes absents et des connaissances de contrôle ou de raisonnement. Le formalisme de règles SWRL (I. Horrocks et al., 2004) est utilisé pour ces cas.
- ❖ des pathologies multi-niveaux (Rodríguez-González, Torres-Niño, et al., 2012) via l'utilisation des propagations transitives (cf. 3.3.4.1) (Seidenberg & Rector, 2006).

L'ontologie obtenue est ensuite instanciée car l'utilisation du formalisme SWRL requiert la présence des instances. A partir de l'ontologie instanciée et du moteur d'inférences JESS (Friedman-Hill, 2008) nous produisons des ensembles (par exemple {signe, maladie, relation, probabilités, etc.}). Ces ensembles sont ensuite stockés dans une base de données MYSQL pour les tests et les évaluations des méthodes via le système ODES (cf. 4.2.9). ODES fournit à partir des ensembles générés précédemment une base de données contenant les relations maladies-signes ainsi que les probabilités associées. Une base de cas permettant de stocker les décisions des praticiens sert de base de données locale (Seroussi et al., 2013) utile pour produire les probabilités propres au site. Un formulaire est généré par le système sur la base du contenu de l'ontologie pour le recueil des signes d'un patient donné et un moteur d'inférences (abductif et probabiliste) est utilisé pour proposer les diagnostics sur la base des signes observés.

#### 4.2.1 L'opérationnalisation des modèles ontologiques

La méthodologie qui a été suivie pour l'élaboration des méthodes de modélisation des signes dans les ontologies biomédicales et des mécanismes de raisonnement associés pour l'aide au diagnostic s'appuie sur la méthode d'opérationnalisation d'une ontologie à quatre niveaux proposée par Peelen (Peelen et al., 2007).

Les connaissances des trois domaines d'application étant toutes bien définies (définition claire, précise, sans ambiguïté) dans la littérature, nous n'avons pas eu recours à l'utilisation de la logique floue (Zadeh, 1965) ou de la « *Fuzzy DL* » (Bobillo & Straccia, 2008; Fernando Bobillo, 2008) pour exprimer certaines connaissances (cf. 3.3.4.3).

Pour illustrer ces méthodes nous utilisons l'exemple de la gammopathie monoclonale à signification indéterminée (MGUS) et du myélome multiple (MM).

Pour rappel une MGUS est caractérisée par une quantité d'immunoglobuline de type A  $< 3\text{g/dL}$ , un taux de plasmocytes dans la moelle osseuse  $< 10\%$  et une absence de CRAB (hypercalcémie, insuffisance rénale, anémie et lésion osseuse) (Grosbois, Decaux, Guenet, Goasguen, & Jegou, 2009; Rajkumar, 2011). Une hypercalcémie est caractérisée par un taux de calcium sérique supérieur à  $2,6\text{ mmol/L}$  (Duron, 1999). Une anémie est caractérisée par un

taux d'hémoglobine dans le sang inférieur à 10 g/dL (de Besoist et al., 2008). Une MGUS a une prévalence de 5% pour les patients âgés de 70 ans et plus. La mesure de l'immunoglobuline de type A < 3 g/dL a une sensibilité de 25% et une spécificité de 90% pour une MGUS (Kyle et al., 2011; Watanaboonyongcharoen et al., 2012).

Le MM est caractérisé par une quantité d'immunoglobuline de type A  $\geq 3$  g/dL, un taux de plasmocytes dans la moelle osseuse  $\geq 10\%$  et une présence de CRAB (Rajkumar, 2011).

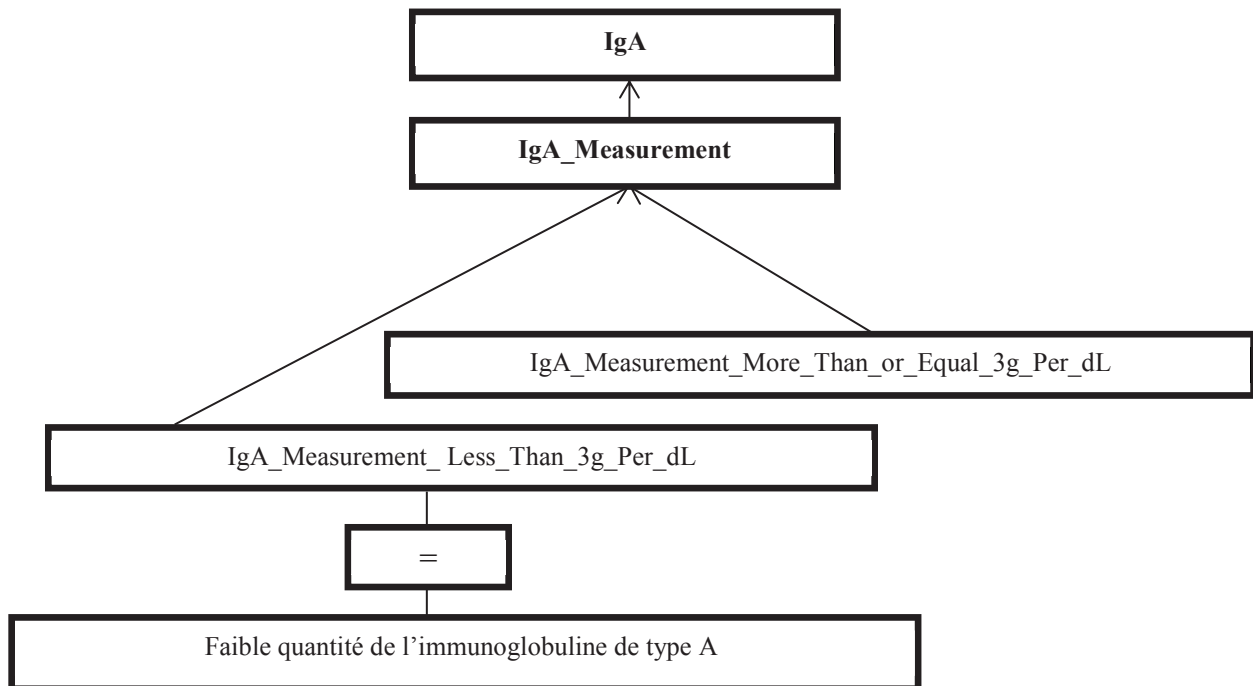
Pour la clarté des schémas, les termes et abréviations du Tableau 6 seront utilisés. Sur tous les schémas le sens de la lecture est celui des flèches. Par exemple, sur la Figure 19 on a une « MGUS » a pour signe « l'immunoglobuline de type A avec une quantité inférieure à 3g/dL ».

**Tableau 6 : Abréviations pour les schémas**

| Nom long  | Terme ou Abréviation  |
|---|---|
| Immunoglobuline de type A   | IgA   |
| Quantité Immunoglobuline de type A  | IgA_Measurement   |
| Quantité Immunoglobuline de type A < 3g/dL                                    | IgA_Measurement_Less_Than_3g_Per_dL   |
| Quantité Immunoglobuline de type A $\geq 3$ g/dL                              | IgA_Measurement_More_Than_Or_Equal_3g_Per_dL                                      |
| Hémoglobine < 10g/dL  | HM_Measurement_Less_Than_10g_Per_dL   |
| Taux de Calcium sérique > 2,6 mmol/L,<br>Taux de Calcium sérique < 2,6 mmol/L | SC_Measurement_More_Than_2_6mmol_Per_L,<br>SC_Measurement_Less_Than_2_6mmol_Per_L |
| Taux de Plasmocyte dans la moelle osseuse                                     | CBM   |
| Gammapathie Monoclonale à signification indéterminée                          | MGUS  |
| Myélome Multiple  | MM  |

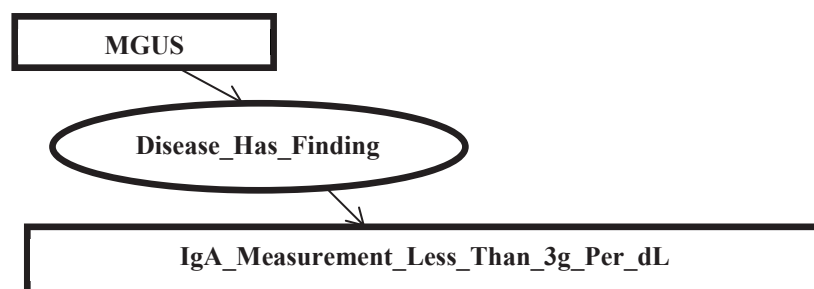
#### 4.2.1.1 Les connaissances dichotomiques (bivaluées)

Dans ce contexte toutes les connaissances quantitatives sont transformées en connaissances qualitatives via la discrétisation. La discrétisation consiste à réaliser un découpage des connaissances quantitatives en classes. Alors, une connaissance comme « faible quantité de l'immunoglobuline de type A » devient « Quantité de l'immunoglobuline de type A < 3 g/dL » pour être conforme aux données de la littérature (Rajkumar, 2011). On peut étendre cela au concept « IgA\_Measurement » qui via une spécialisation est découpée en deux classes (Figure 18) : IgA\_Measurement\_Less\_Than\_3g\_Per\_dL (condition pour une MGUS) et IgA\_Measurement\_More\_Than\_or\_Equal\_3g\_Per\_dL (condition pour un MM).



**Figure 18:** Spécialisation du concept `IgA_Measurement` en utilisant les données de la littérature. Les deux spécialisations sont respectivement des conditions de MM et MGUS.

Deux étapes sont nécessaires, une première permettant de mieux exprimer le concept en caractérisant la valeur seuil et la deuxième étape consistant à créer des classes. A partir de ce découpage il est possible de représenter la restriction suivante (Figure 19) : une « MGUS » a pour signe « `IgA_Measurement_Less_Than_3g_Per_dL` ». Cette représentation se fait par le biais des relations présentées dans le tableau 3.



**Figure 19:** Représentation des connaissances bivaluées. Expression de la restriction : une MGUS a pour signe le « `IgA_Measurement_Less_Than_3g_Per_dL` ».

De manière analogue en opérant par discrétisation les connaissances taux de calcium sérique inférieur à 2,6 mmol/L, taux d'hémoglobine inférieur à 10g/dL se présenteraient comme ci-après :

```

Serum_calcium_Measurement
  SC_Measurement_More_Or_Equal_Than_2_6mmol_Per_L
  SC_Measurement_Less_Than_2_6mmol_Per_L
Hemoglobin_Measurement
  HM_Measurement_Less_Than_10g_Per_dL
  
```

### 4.2.1.2 Les connaissances multivaluées

D'autres connaissances sont multivaluées, c'est-à-dire qu'elles peuvent prendre plusieurs valeurs. Par exemple : la quantité d'immunoglobuline de type A (quantitative continue), le nombre de lésions osseuses (quantitatif discret), etc. Certaines de ces connaissances nécessitent d'être caractérisées par une valeur numérique accompagnée d'une unité. Par exemple, la quantité d'immunoglobuline de type A est exprimée en gramme par litre (g/L) ou gramme par décilitre (g/dL) (Figure 20).

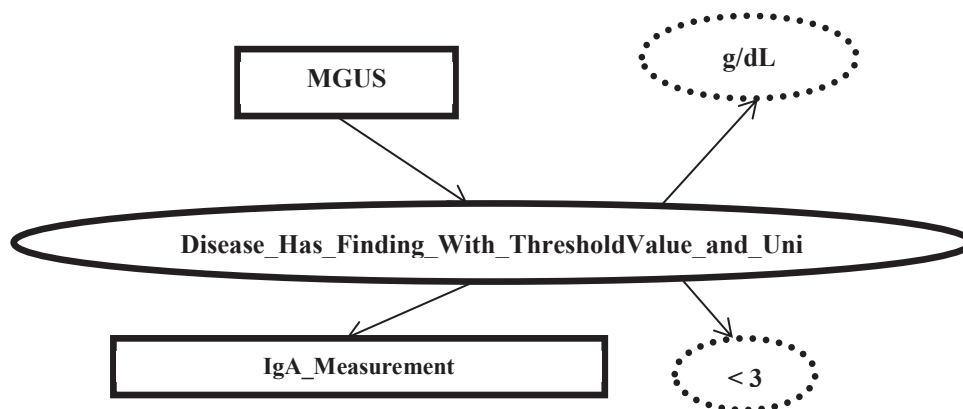


Figure 20: Expression de la restriction : une MGUS a pour signe le « IgA\_Measurement < 3g/dL ». Cette restriction n'est pas exprimable directement en OWL car la relation « Disease\_Has\_Finding\_With\_ThresholdValue\_and\_Unit » est d'arité 4 ou d'arité n.

Des relations comme celles présentées dans la Figure 20 ne peuvent pas être directement représentées en OWL (Stevens et al., 2007), pour cela nous devons utiliser l'ontology design pattern 1 (ODP 1) du W3C (Hayes et al., 2006) (cf. 3.3.2.4). L'ODP 1 permet d'exprimer ce type de relations par des classes via la réification.

Une nouvelle classe « *Disease\_Finding* » est créée pour faire office de superclasse pour toutes les classes représentant les couples (maladie, signe) issues de la réification des relations d'arité n. Par exemple, le couple (MGUS, IgA\_Measurement) est « *MGUS\_IgA\_Measurement* ». Quatre relations d'arité 2 (Tableau 7) dont deux ObjectProperties (Has\_Disease, Has\_Finding) et deux DatatypeProperties (Has\_Unit, Has\_Value) sont créées et exploitées pour représenter ces connaissances. Pour rappel, les relations de type « *ObjectProperty* » permettent de lier une classe à une autre classe et les relations de type « *DatatypeProperty* » lient une classe à une valeur (Cardoso, 2007; Malviya et al., 2011) (cf. 3.3.2.5). Ces relations sont les suivantes :

- ❖ « Has\_Disease » qui lie la nouvelle classe créée représentant le couple (maladie, signe) à la classe de la maladie.
- ❖ « Has\_Finding » lie le couple (maladie, signe) au signe.
- ❖ « Has\_Unit » qui lie le couple (maladie, signe) à l'unité.
- ❖ « Has\_Value » qui lie le couple (maladie, signe) à une valeur qui représente le seuil.

A ces nouvelles relations s'ajoute la relation « May\_Have\_Finding » qui lie le couple (maladie, signe) au signe dans certains cas.



Tableau 7: Tableau des nouvelles relations de l'ontologie opérationnalisée.

| Nom              | Domaine         | Range                  | Arité |
|------------------|-----------------|------------------------|-------|
| Has_Disease      | Disease_Finding | Diseases_And_Disorders | 2     |
| Has_Finding      | Disease_Finding | Finding                | 2     |
| May_Have_Finding | Disease_Finding | Finding                | 2     |
| Has_Unit         | Disease_Finding | String                 | 2     |
| Has_Value        | Disease_Finding | Integer, float, String | 2     |

Cette structure permet de représenter en OWL le fait qu'une **MGUS** a pour signe « IgA\_Measurement < **3g/dL** » (Rajkumar, 2011).

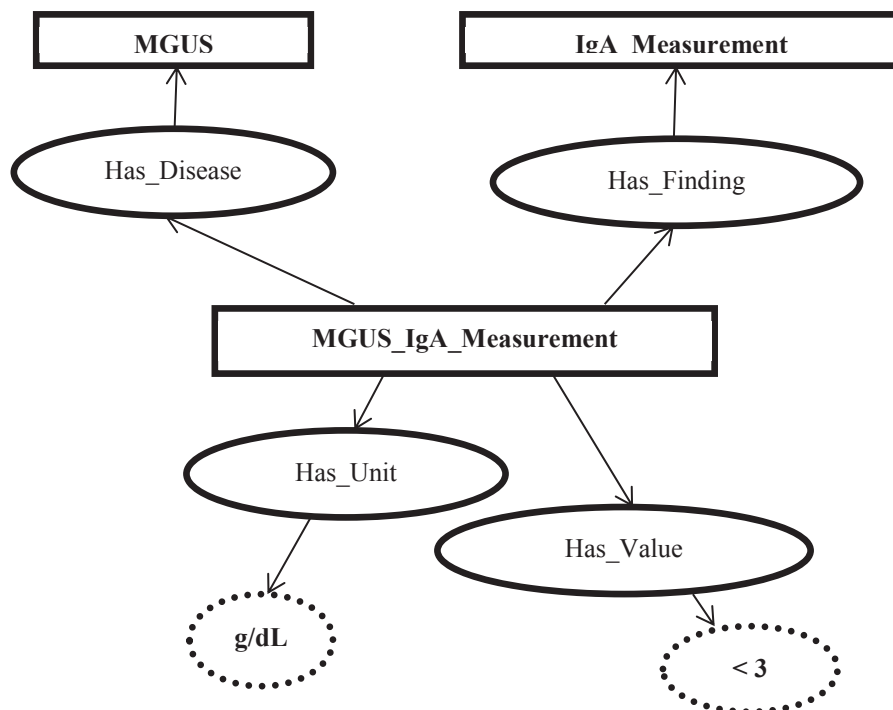


Figure 21: Expression des connaissances multivaluées. Nous avons réifié la relation d'arité 4 en créant une classe « **MGUS\_IgA\_Measurement** » représentant le couple (MGUS,IgA\_Measurement). A cette nouvelle classe nous lions le signe IgA\_Measurement, la valeur de la mesure du signe et l'unité. Toutes les nouvelles relations créées pour exprimer ces restrictions sont d'arité 2. Pour finir, nous lions la classe « **MGUS** » à notre classe « **MGUS\_IgA\_Measurement** ».

Pour les illustrations qui vont suivre nous utiliserons les représentations dichotomiques de la mesure de l'immunoglobuline « **IgA\_Measurement\_Less\_Than\_3g\_Per\_dL** ou **IgA\_Measurement\_More\_Than\_or\_Equal\_3g\_Per\_dL** ».

#### 4.2.1.3 Les connaissances probabilistes

Nous proposons de représenter la sémiologie quantitative dans les ontologies (Chatellier et al., 1996). Ces connaissances sont utiles car elles permettent via des méthodes probabilistes comme la méthode bayésienne de dériver les probabilités d'avoir une maladie ou pas. Ces connaissances permettent d'exprimer par exemple, le fait qu'une MGUS a pour signe

l'immunoglobuline avec une quantité inférieure à 3g/dL avec une sensibilité de 25% et une spécificité de 90%. Et aussi, le fait qu'une MGUS a une prévalence de 5% pour les patients ayant plus de 70 ans. Les sensibilités et spécificités présentent l'intérêt ne pas être trop influencées par le contexte (Chatellier et al., 1996). En revanche la prévalence est fonction du contexte. Dans notre exemple, le contexte est celui « des patients ayant plus de 70 ans ». Trois relations d'arité 3 sont nécessaires pour représenter ces connaissances :

- ❖ « Disease\_Has\_Finding\_Se » qui lie une MGUS au signe « IgA\_Measurement\_Less\_Than\_3g\_Per\_dL » avec une sensibilité de 25% soit 0,25.
- ❖ « Disease\_Has\_Finding\_Sp » qui lie une MGUS au signe « IgA\_Measurement\_Less\_Than\_3g\_Per\_dL » avec une sensibilité de 90% soit 0,9.
- ❖ « Disease\_Prevalence\_Context » qui lie une MGUS au contexte avec une prévalence de 5% soit 0,05.

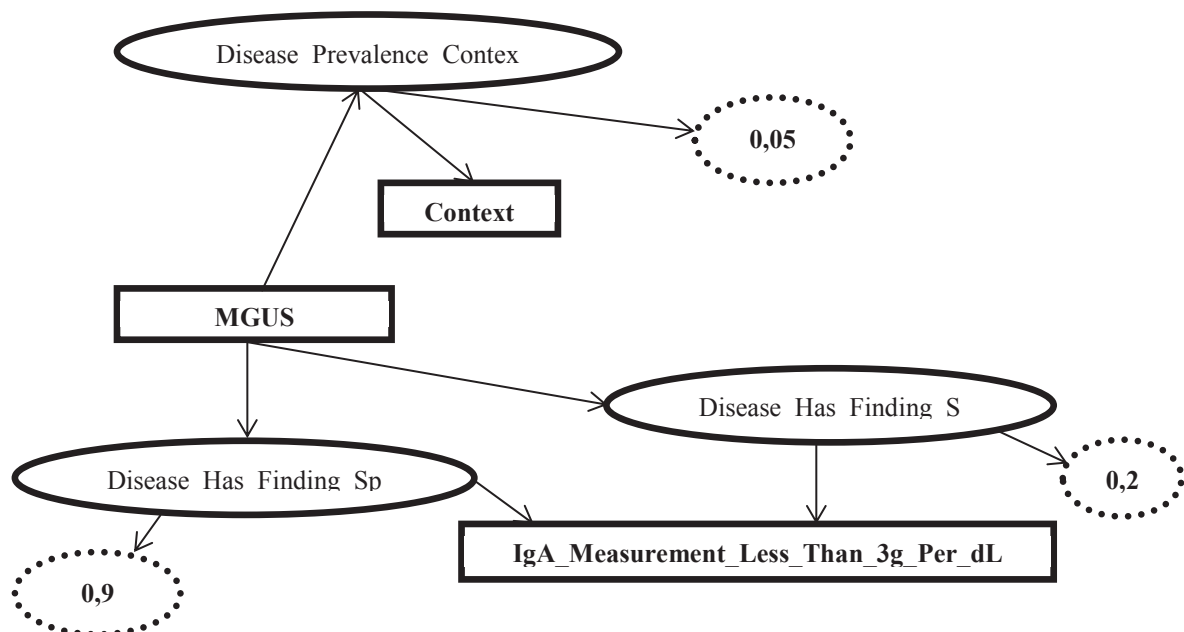


Figure 22: Expression des connaissances probabilistes. Représentation de la restriction : une MGUS a une prévalence de 5% pour les patients de plus de 70 ans et a pour signe le « IgA\_Measurement < 3g/dL » avec une sensibilité de 25% et une spécificité de 90%. Trois relations d'arité 3 « Disease\_Prevalence\_Context, Disease\_Has\_Finding\_Se, Disease\_Has\_Finding\_Sp » sont nécessaires pour cette représentation. Ces trois relations ne sont pas exprimables comme telles par le langage OWL. Pour ce faire nous devons utiliser la réification comme ci-après à la Figure 23.

Comme pour les cas précédents six relations d'arité 2 (Tableau 8) dont trois ObjectProperties (Has\_Context, Has\_Disease, Has\_Finding) et trois DatatypeProperties (Has\_Sensitivity, Has\_Specificity, Has\_Prevalence) sont créées et exploitées pour représenter ces connaissances. Elles sont les suivantes :

- ❖ « Has\_Context » qui lie le couple (maladie, contexte) au contexte,
- ❖ « Has\_Sensitivity » qui lie le couple (maladie, signe) à une valeur numérique,
- ❖ « Has\_Specificity » qui lie le couple (maladie, signe) à une valeur numérique,
- ❖ « Has\_Prevalence » qui lie le couple (maladie, contexte) à une valeur numérique.

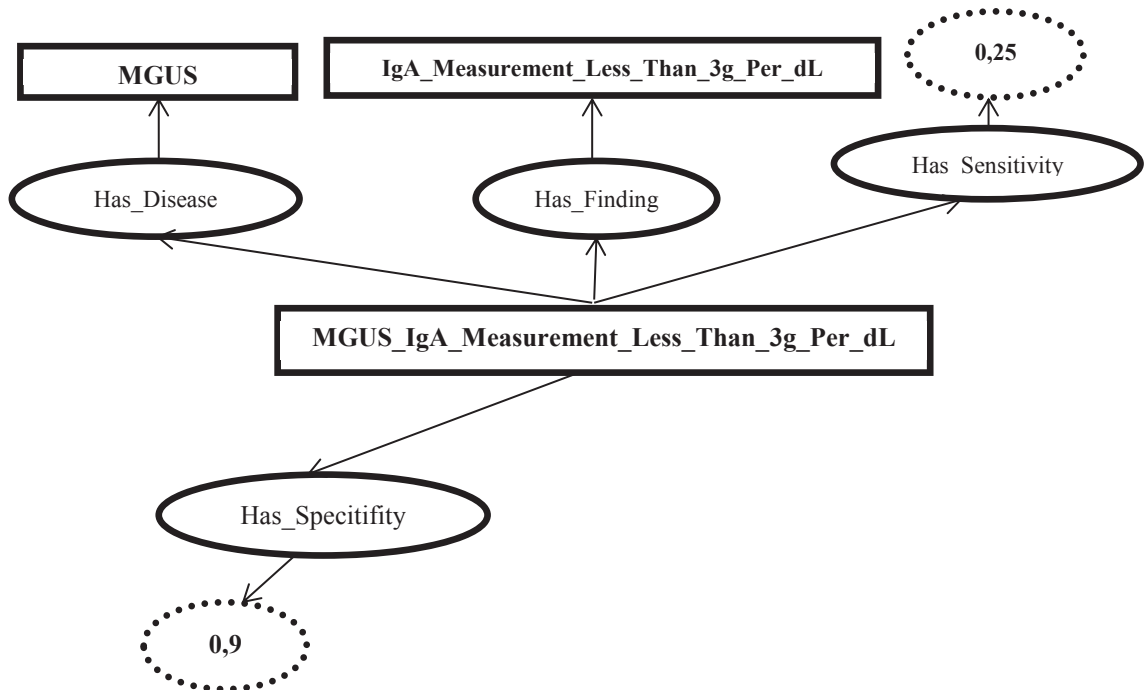
Une nouvelle classe « Disease\_Context » est créée pour faire office de superclasse pour toutes les classes représentant les couples (maladie, contexte) issues de la réification de la relation d'arité 3 « Disease\_Prevalence\_Context ». Par exemple, le couple (MGUS, Contexte) est

« *MGUS\_Context* ». Les relations « Has\_Disease et Has\_Finding » ont déjà été présentées dans la sous-section précédente (cf. 4.2.1.2).

**Tableau 8: Tableau des nouvelles relations utiles pour la représentation des probabilités dans l'ontologie.**

| Nom             | Domaine         | Range   | Arité |
|-----------------|-----------------|---------|-------|
| Has_Sensitivity | Disease_Finding | Float   | 2     |
| Has_Specificity | Disease_Finding | Float   | 2     |
| Has_Prevalence  | Disease_Context | Float   | 2     |
| Has_Context     | Disease_Context | Context | 2     |

Cette structure permet de représenter en OWL le fait qu'une MGUS a pour signe «IgA\_Measurement\_Less\_Than\_3g\_Per\_dL » avec une sensibilité de 25% et une spécificité de 90%.



**Figure 23: Expression des sensibilités et spécificités des signes pour les maladies. Représentation de la restriction : une MGUS a pour signe le « IgA\_Measurement\_Less\_Than\_3g\_Per\_dL » avec une sensibilité de 25% et une spécificité de 90%.**

La structure suivante (Figure 24) permet de représenter en OWL le fait qu'une MGUS a une prévalence de 5% pour une population de personnes âgées de plus de 70 ans. La population des personnes âgées de plus de 70 ans est matérialisée dans la Figure 24 par la classe « Context ».

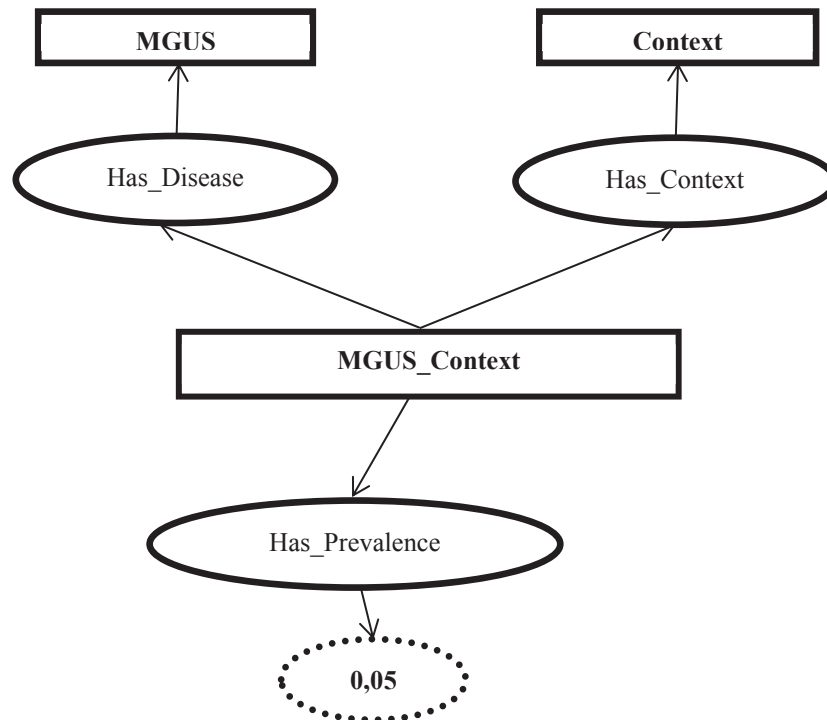


Figure 24 : Expression de la prévalence d'une maladie pour une population donnée. La population donnée dans notre exemple est celle des patients de plus de 70 ans et elle est matérialisée par la classe « Context ». Pour cela nous avons créé une classe représentant le couple (Maladie, Contexte) soit « MGUS\_Context ». Deux relations d'arité 2 dont une de type DatatypeProperty « Has\_Prevalence » et une de type ObjectProperty « Has\_Context » ont été ajoutées.

#### 4.2.1.4 Les définitions multi-niveaux

Les définitions multi-niveaux concernent les cas où certaines maladies contiennent d'autres maladies dans leurs définitions (Rodríguez-González, Torres-Niño, et al., 2012). On considère que DHF est la relation « *Disease\_Has\_Finding* » (Tableau 3).

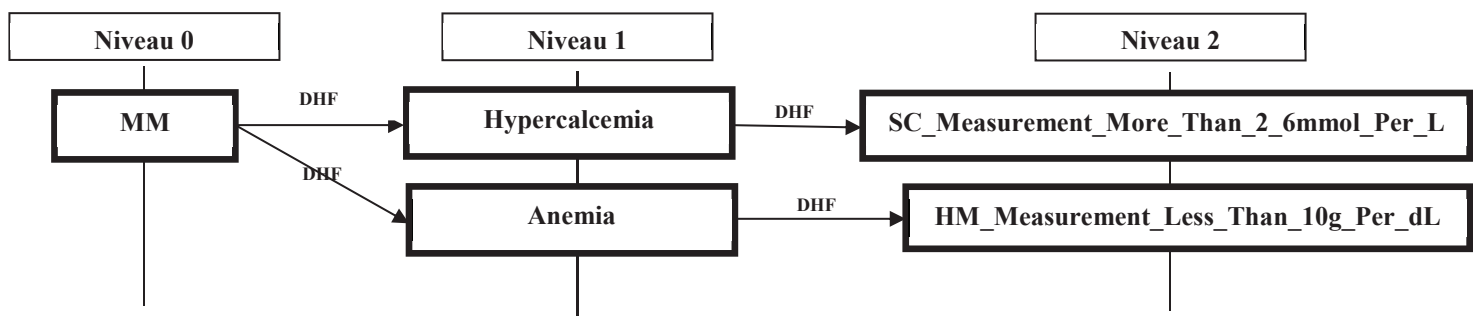


Figure 25: Définition du MM sous forme de maladie multi-niveau. Le MM est représenté ici sous forme de pathologie de niveau ou profondeur 2. L'hypercalcémie y est définie via la mesure du calcium sérique et l'anémie y est définie via la mesure de l'hémoglobine.

Notre proposition pour la gestion des définitions multi-niveaux se fonde sur l'utilisation de la propagation transitive (Seidenberg & Rector, 2006). En effet, si le MM « a pour signe » l'hypercalcémie, et l'hypercalcémie « a pour signe » un taux de calcium sérique supérieur à 2,6 mmol/L, alors le MM « a pour signe » un taux de calcium sérique supérieur à 2,6 mmol/L. Cette propagation transitive, nous la matérialisons par le chaînage des relations (cf. 3.3.2.4) et nous la mettons en œuvre via des règles SWRL (Règle 9 et 10 cf. 4.2.3) (I. Horrocks et al., 2004).

### 4.2.2 La représentation des signes absents ou négatif

La gestion des signes absents ou négatifs est faite en utilisant les relations de types « Excludes ». En effet, notre approche de représentation s'appuie sur cette relation de type « Excludes » définie par les ingénieurs du NCIT (Golbeck et al., 2003), et elle est similaire à l'approche de Werner Ceusters (Ceusters et al., 2006, 2007) (cf. 3.3.4.5). Le signe absent ou négatif ne se traduit pas en termes de concepts (classe au sens de PROTÉGÉ), mais en termes de relations entre une maladie M et un signe S.

### 4.2.3 Le raisonnement abductif

Nous avons précédemment proposé des solutions de modélisation pour les : (a) connaissances dichotomiques, (b) connaissances multivaluées, (c) connaissances probabilistes, (d) définitions multi-niveaux et, (e) signes absents ou négatifs. Notre objectif dans cette section est de proposer une méthode pour mettre en œuvre le diagnostic médical (cf. 2.3), en d'autres termes le raisonnement abductif (cf. 2.3.3).

Plusieurs travaux montrent que les moteurs d'inférences mettant en œuvre OWL classique ne permettent pas de faire du raisonnement médical (cf. 3.3.3, 3.3.4.1, 3.3.4.6). Notre proposition s'appuie sur les travaux de Cory Henson (cf. 3.3.4.6) (Henson et al., 2011). En effet, notre approche pour mettre en œuvre le raisonnement abductif s'inspire des « Parsimonious Covering Theory » (PCT) de Cory Henson. Nous construisons un réseau de relations entre les signes et les maladies (cf. 3.3.4.6). Pour ce faire, nous intégrons de nouvelles connaissances dans le modèle ontologique, à savoir les connaissances de contrôle via des règles. Ces règles sont créées sur la base des stratégies de raisonnement des experts et des spécialistes en médecine lors du processus diagnostique (Arocha et al., 2005). La création de ces règles impose l'ajout de nouvelles relations d'une part (Tableau 9), et d'autre part la création de nouvelles variables. Ces ajouts permettent la création des ensembles de règles qui sont utilisés par le moteur d'inférences mis en place.

**Tableau 9: Tableau des nouvelles relations utiles pour les règles ajoutées dans l'ontologie opérationnalisée.**

| Nom                                | Domaine | Range                                 | Arité  |
|------------------------------------|---------|---------------------------------------|--------|
| Finding_Has_Diagnosis              | Finding | Diseases_And_Disorders, Float, String | 2 ou 4 |
| Finding_Excludes_Diagnosis         | Finding | Diseases_And_Disorders, Float, String | 2 ou 4 |
| Finding_Absence_Excludes_Diagnosis | Finding | Diseases_And_Disorders, Float, String | 2 ou 4 |
| Finding_Absence_Has_Diagnosis      | Finding | Diseases_And_Disorders, Float, String | 2 ou 4 |

Pour les règles qui sont présentées dans cette section, on considère d et d1 comme des « *Diseases\_And\_Disorders* ». On prend f comme un « *Finding* » qui peut être à caractère dichotomique ou multivalué. On ajoute les variables df comme une « *Disease\_Finding* », tv une valeur seuil (chaîne, entière, réelle, ect.) et u l'unité (chaîne). Les règles 1,3,5 et 7 concernent les cas où nous avons des connaissances dichotomiques (signes bivalents). Plus loin, ces règles sont utilisées pour créer les ensembles de triplets (cf. 4.2.7). Les règles 2,4,6 et

8 concernent les cas où nous avons des connaissances multivaluées (signes à valeurs seuils). Celles-ci sont utilisées pour créer les ensembles de quintuplets (cf. 4.2.7). Les règles 9 et 10 concernent les cas de diagnostic multi-niveaux (Rodríguez-González, Torres-Niño, et al., 2012). Celles-ci sont aussi utilisées pour créer les ensembles de triplets (cf. 4.2.7).

(a) Si le signe  $f$  est présent et ce signe  $f$  est une manifestation possible de la maladie  $d$ , alors le signe  $f$  est un signe possible pour le diagnostic  $d$

$$Disease\_May\_Have\_Finding(d,f) \wedge Finding(f) \wedge Diseases\_And\_Disorders(d) \rightarrow Finding\_Has\_Diagnosis(f,d). \quad (R\grave{e}gle.1)$$

$$Has\_Finding(df,f) \wedge Finding(f) \wedge Disease\_Finding(df) \wedge Has\_Disease(df,d) \wedge Diseases\_And\_Disorders(d) \wedge Has\_Value(df,tv) \wedge Has\_Unit(df,u) \rightarrow Finding\_Has\_Diagnosis(f,d,tv,u). \quad (R\grave{e}gle.2)$$

(b) Si le signe  $f$  est présent et ce signe  $f$  exclut la maladie  $d$ , alors le signe  $f$  exclut l'hypothèse  $d$  comme diagnostic possible

$$Disease\_Excludes\_Finding(d,f) \wedge Finding(f) \wedge Diseases\_And\_Disorders(d) \rightarrow Finding\_Excludes\_Diagnosis(f,d). \quad (R\grave{e}gle.3)$$

$$Excludes\_Finding(df,f) \wedge Finding(f) \wedge Disease\_Finding(df) \wedge Has\_Disease(df,d) \wedge Diseases\_And\_Disorders(d) \wedge Has\_Value(df,tv) \wedge Has\_Unit(df,u) \rightarrow Finding\_Excludes\_Diagnosis(f,d,tv,u). \quad (R\grave{e}gle.4)$$

(c) Si le signe  $f$  est absent et ce signe  $f$  est requis pour la maladie  $d$ , alors l'absence du signe  $f$  exclut l'hypothèse  $d$  comme diagnostic possible

$$Disease\_Has\_Finding(d,f) \wedge Finding(f) \wedge Diseases\_And\_Disorders(d) \rightarrow Finding\_Absence\_Excludes\_Diagnosis(f,d). \quad (R\grave{e}gle.5)$$

$$Has\_Finding(df,f) \wedge Finding(f) \wedge Disease\_Finding(df) \wedge Has\_Disease(df,d) \wedge Diseases\_And\_Disorders(d) \wedge Has\_Value(df,tv) \wedge Has\_Unit(df,u) \rightarrow Finding\_Absence\_Excludes\_Diagnosis(f,d,tv,u). \quad (R\grave{e}gle.6)$$

(d) Si le signe  $f$  est absent et la présence de ce signe  $f$  exclut la maladie  $d$  alors l'absence de ce signe rend l'hypothèse  $d$  possible

$$Disease\_Excludes\_Finding(d,f) \wedge Finding(f) \wedge Diseases\_And\_Disorders(d) \rightarrow Finding\_Absence\_Has\_Diagnosis(f,d). \quad (R\grave{e}gle.7)$$

$$Excludes\_Finding(df,f) \wedge Finding(f) \wedge Disease\_Finding(df) \wedge Has\_Disease(df,d) \wedge Diseases\_And\_Disorders(d) \wedge Has\_Value(df,tv) \wedge Has\_Unit(df,u) \rightarrow Finding\_Absence\_Has\_Diagnosis(f,d,tv,u). \quad (R\grave{e}gle.8)$$

(e) Si  $f$  est un signe pour la maladie  $d$  et la maladie  $d$  est un signe pour la maladie  $d1$  alors  $f$  est un signe qui rend l'hypothèse  $d1$  possible

$$\begin{aligned} & Disease\_Has\_Finding(d,f) \wedge Disease\_Has\_Finding(d1,d) \wedge Finding(f) \wedge \\ & Diseases\_And\_Disorders(d) \wedge Diseases\_And\_Disorders(d1) \rightarrow Finding\_Has\_Diagnosis \\ & (f,d1). \end{aligned} \quad (R\grave{e}gle.9)$$

(f) Si f est un signe pour la maladie d et la maladie d1 exclut la maladie d alors le signe f exclut le diagnostic d1

$$\begin{aligned} & Disease\_Has\_Finding(d,f) \wedge Disease\_Excludes\_Finding(d1,d) \wedge Finding(f) \wedge \\ & Diseases\_And\_Disorders(d) \wedge Diseases\_And\_Disorders(d1) \rightarrow Finding\_Excludes\_Diagnosis \\ & (f,d1). \end{aligned} \quad (R\grave{e}gle.10)$$

#### 4.2.4 Le raisonnement probabiliste

Les règles précédentes permettent de prendre en charge le raisonnement abductif. Nous proposons maintenant de réaliser des inférences bayésiennes. Ce mode d'inférence permet de dériver les probabilités *a posteriori* en nous appuyant sur les sensibilités (SE) et spécificités (SP) des signes pour les maladies d'une part, et d'autre part, des prévalences (Prev) des maladies pour une population donnée et des règles dites de Bayes (McGee, 2002)(Attia, 2003) (cf. 2.4). Notre choix s'appuie sur la méthode de calcul prenant en compte les rapports de vraisemblance car elle permet de prendre en compte la combinaison des observations (signes présents et signes absents).

Deux nouvelles règles sont nécessaires pour exprimer les rapports de vraisemblance dans l'ontologie opérationnalisée. Les relations « *Finding\_Diagnosis\_LrPlus* », « *Finding\_Diagnosis\_LrMinus* » sont nécessaires pour ces nouvelles règles. Dans ces règles (Règle. 11 et Règle. 12), d est une hypothèse de maladie, f est un signe clinique ou symptôme, df est un couple (maladie, signe), se est une valeur de sensibilité, sp est une valeur de spécificité, dc est un couple (maladie, contexte), c est un contexte et prev1 une valeur de prévalence.

$$\begin{aligned} & (g) Has\_Finding(df,f) \wedge Disease\_Finding(df) \wedge Finding(f) \wedge Has\_Sensitivity(df,se) \wedge Has\_Specificity \\ & (df,sp) \wedge Has\_Disease(df,d) \wedge Has\_Prevalence(dc,prev1) \wedge Has\_Context(dc,c) \wedge Context(c) \wedge Disease\_Context(dc) \wedge Has\_Disease(dc,d) \wedge Diseases\_And\_Disorders(d) \rightarrow \\ & Finding\_Diagnosis\_Lrplus(f,d,c,prev1,se, sp,se/(1-sp)). \end{aligned} \quad (R\grave{e}gle.11)$$

$$\begin{aligned} & (h) Has\_Finding(df,f) \wedge Disease\_Finding(df) \wedge Finding(f) \wedge Disease\_Finding\_Se(df,se) \wedge Disease\_Finding\_Sp(df,sp) \wedge Has\_Disease(df,d) \wedge Has\_Prevalence(dc,prev1) \wedge Has\_Context(dc,c) \wedge Context(c) \wedge Disease\_Context(dc) \wedge Has\_Disease(dc,d) \wedge Diseases\_And\_Disorders(d) \rightarrow \\ & Finding\_Diagnosis\_Lrminus(f,d,c,prev1,se,sp,(1-se)/sp). \end{aligned} \quad (R\grave{e}gle.12)$$

Les règles 11 et 12 serviront à créer les ensembles de septuplets (cf. 4.2.7).

#### 4.2.5 La création des cas prototypiques

Le formalisme de représentation des règles SWRL et les formalismes d'interrogation SQWRL et SWRL-IQ des ontologies que nous utilisons pour mettre en œuvre les règles présentées précédemment ne fonctionnent qu'avec des instances (Motik, Sattler, & Studer, 2005). Pour cela, les modèles doivent être instanciés. L'instanciation dans ce contexte consiste à créer des



instances (individus) pour chaque classe de l'ontologie (maladies et signes). Par exemple pour une MGUS, il s'agit de créer une instance pour la classe représentant une MGUS, une instance pour chaque classe représentant ses signes et de lier l'instance de la classe MGUS à celles de ses signes via leurs relations.

#### 4.2.5.1 La base de règles d'ordre 0

C'est une retranscription totale de la Tbox en Abox qui permet à partir des règles d'ordre 1 (cf. 4.2.3) créées précédemment de générer un ensemble de règles d'ordre 0 qui peuvent être manipulées via des outils comme SQWRL et SWRL-IQ. Cette retranscription nous permet d'instancier notre réseau de relations maladie-signes (cf. 3.3.4.6 et 4.2.3).

#### 4.2.5.2 La prise en charge de la hiérarchie des classes

Dans ce processus d'instanciation, la hiérarchie des classes est conservée. Par exemple, le myélome multiple (MM) « a pour signe » les lésions osseuses (LO) (Rajkumar, 2011), et les LO ont des sous-types de lésions : (a) la lésion lytique (LT), (b) la lésion métastatique lytique (LM), etc. Lors de l'instanciation, une instance de MM est créée, les instances respectives de LO, LT et LM sont créées, puis ces instances sont liées au MM via la relation « a pour signe » (Figure 53).

#### 4.2.6 Les axiomes de fermeture (Closure axiom)

Des axiomes de fermeture (Horridge et al., 2004) sont intégrés aux définitions existantes conformément à la procédure décrite dans la sous-section 3.3.2.4.

#### 4.2.7 La création des ensembles

A partir des règles d'ordre 1 ajoutées, des instances créées dans l'ontologie et des APIs (JESS, SWRL, SWRL-IQ) nous produisons des ensembles. En considérant que :

- ❖  $r$  est une relation,
- ❖  $f$  est un signe, un symptôme ou une observation de manière générale,
- ❖  $d$  est une maladie ou une hypothèse de maladie,
- ❖  $c$  est un contexte (par exemple, les hommes âgés de plus de 70 ans),
- ❖  $Se$  est une sensibilité de  $f$  pour  $d$ ,
- ❖  $Sp$  est une spécificité de  $f$  pour  $d$ ,
- ❖  $prev$  est la prévalence de  $d$  selon  $c$ ,
- ❖  $LrPlus$  est un rapport de vraisemblance positif de  $f$  pour  $d$ ,
- ❖  $LrMinus$  est un rapport de vraisemblance négatif de  $f$  pour  $d$ ,
- ❖  $tv$  est la valeur seuil de  $f$  pour  $d$ , et
- ❖  $u$  est l'unité de  $f$  pour  $d$  pour la valeur seuil  $tv$ .

A partir de ces variables nous avons des ensembles de :

- ❖ triplet  $(r, f, d)$ ,
- ❖ quintuplet  $(r, f, d, tv, u)$  et
- ❖ septuplets  $(f, d, c, prev, Se, Sp, LrPlus)$  et  $(f, d, c, prev, Se, Sp, LrMinus)$ .

Ces ensembles sont stockés dans une base de données relationnelle MYSQL pour les tests et évaluations des modèles élaborés.

### 4.2.8 Le moteur d'inférences

Le moteur d'inférences, sur la base des ensembles stockés dans la base de données MYSQL et des observations faites par le praticien extrait dans un premier temps via le raisonnement abductif (Vertue & Haig, 2008; Walton, 2004) les hypothèses de maladies, puis il trie ces hypothèses selon deux modes et de manière exclusive : le calcul des ratios ou l'inférence bayésienne.

Le mode de raisonnement abductif exploite les triplets (si les paramètres sont dichotomiques, c'est-à-dire présent ou absent) ou les quintuplets (si les paramètres sont multivalués) pour extraire les hypothèses des maladies. Si les probabilités sont toutes disponibles, le mode de raisonnement probabiliste trie les hypothèses de maladies extraites à la première étape en s'appuyant sur les septuplets. Sinon, si les probabilités ne sont pas toutes disponibles, les hypothèses de maladies sont triées sur la base d'un ratio (nombre de signes observés impliquant chaque hypothèse de maladie / nombre de signes de chaque hypothèse de maladie) (Figure 59).

### 4.2.9 Architecture et fonctionnement de l'Ontology Driven Expert System (ODES)

Cette section a pour but de présenter l'architecture logique et le fonctionnement du système ODES. ODES est un système expert d'aide au diagnostic médical s'appuyant sur une ontologie sémiologique biomédicale. ODES est fondé sur une architecture client-serveur classique. Il possède (a) une base de connaissances contenant toutes les informations sur les maladies et les signes, (b) une base de cas contenant toutes les décisions des praticiens (utilisateurs), et une interface qui contient les paramètres ou les signes et les symptômes des patients. Ces paramètres sont notifiés par les utilisateurs et sont utilisés par le système pour inférer les hypothèses de maladies à partir de la base de connaissances. Les paramètres peuvent être de types quantitatif (par exemple binaire, numérique), qualitatif (par exemple ordinale, nominale). Les termes paramètres, signes et symptômes dans le cadre du système informatique sont utilisés de manière interchangeable.

#### 4.2.9.1 Les cas d'utilisation

Un cas d'utilisation représente une unité d'interaction entre un utilisateur (humain ou machine) et un système. C'est un ensemble d'actions réalisées par le système, en réponse à une action d'un acteur. Il est une unité significative de travail. Dans un diagramme de cas d'utilisation (Figure 26), les utilisateurs sont appelés acteurs (actors), ils interagissent avec les cas d'utilisation (use cases). Les acteurs du système sont les suivants :

- ❖ médecin (tout utilisateur étant du corps médical) est celui qui utilise le système au quotidien pour ces tâches d'aide au diagnostic médical.
- ❖ développeur est celui qui se charge de la vérification du fonctionnement du système et des mises à jour des fonctionnalités.
- ❖ administrateur est celui qui se charge de la vérification de la cohérence et de la qualité des résultats (proposition du système et décision des médecins). Cet administrateur peut être médecin ou non.

Les cas d'utilisation sont les suivants :

- ❖ chargement de l'ontologie : L'ontologie est chargée dans la mémoire du système et envoyée au module d'instanciation.
- ❖ instanciation de l'ontologie : L'ontologie est instanciée. Toutes les maladies qui sont décrites au niveau de la TBox sont instanciées en s'appuyant sur leurs descriptions intensionnelles.
- ❖ chargement du contenu de l'ontologie dans une base de données relationnelle MYSQL : Les ensembles créés (cf. 4.2.7) sont enregistrés dans la base de données MYSQL.
- ❖ connexion : Tous les utilisateurs du système ont un compte et doivent s'authentifier pour pouvoir l'exploiter.
- ❖ sélection de l'ontologie : Le système est multi-ontologique, plusieurs bases de connaissances sont accessibles et, en fonction de la base de connaissances le système génère le formulaire correspondant.
- ❖ mise à jour des probabilités : Elle permet de mettre à jour les sensibilités, spécificités des signes pour les maladies et les prévalences des maladies pour la population considérée (selon les contextes prédéfinis).
- ❖ formulaire adaptatif : Formulaire de saisie des signes ou observations du médecin.
- ❖ notification des observations du médecin : Il s'agit de la saisie des observations du médecin sur le formulaire adaptatif.
- ❖ résultats/ diagnostic / hypothèse de maladie : Résultats renvoyés par le système après la saisie des observations du médecin.
- ❖ enregistrement des décisions – mise à jour de la base de données : Toutes les décisions des médecins sont stockées dans une base de données (base des cas). Ces données sont ensuite utilisées pour calculer les sensibilités et spécificités des signes pour les maladies et les prévalences des maladies pour les populations considérées.
- ❖ vérification des résultats : L'administrateur du système vérifie les résultats renvoyés par le système ODES en permanence afin de déceler des éventuelles incohérences qui, lorsqu'elles sont avérées sont envoyées au développeur qui se charge de les corriger. Cette vérification consiste à chercher les cas où la décision du médecin n'est pas dans la liste des propositions du système.
- ❖ mise à jour de la base de connaissances / apprentissage : Les sensibilités et spécificités des signes pour les maladies et les prévalences des maladies pour les populations considérées selon les contextes spécifiques sont recalculées en permanence par le système en fonction du contenu de la base des cas. Ces valeurs peuvent à tout moment être utilisées pour mettre à jour la base de connaissances via un processus d'apprentissage supervisé. Cette opération est menée par une personne ayant des habilitations spécifiques.

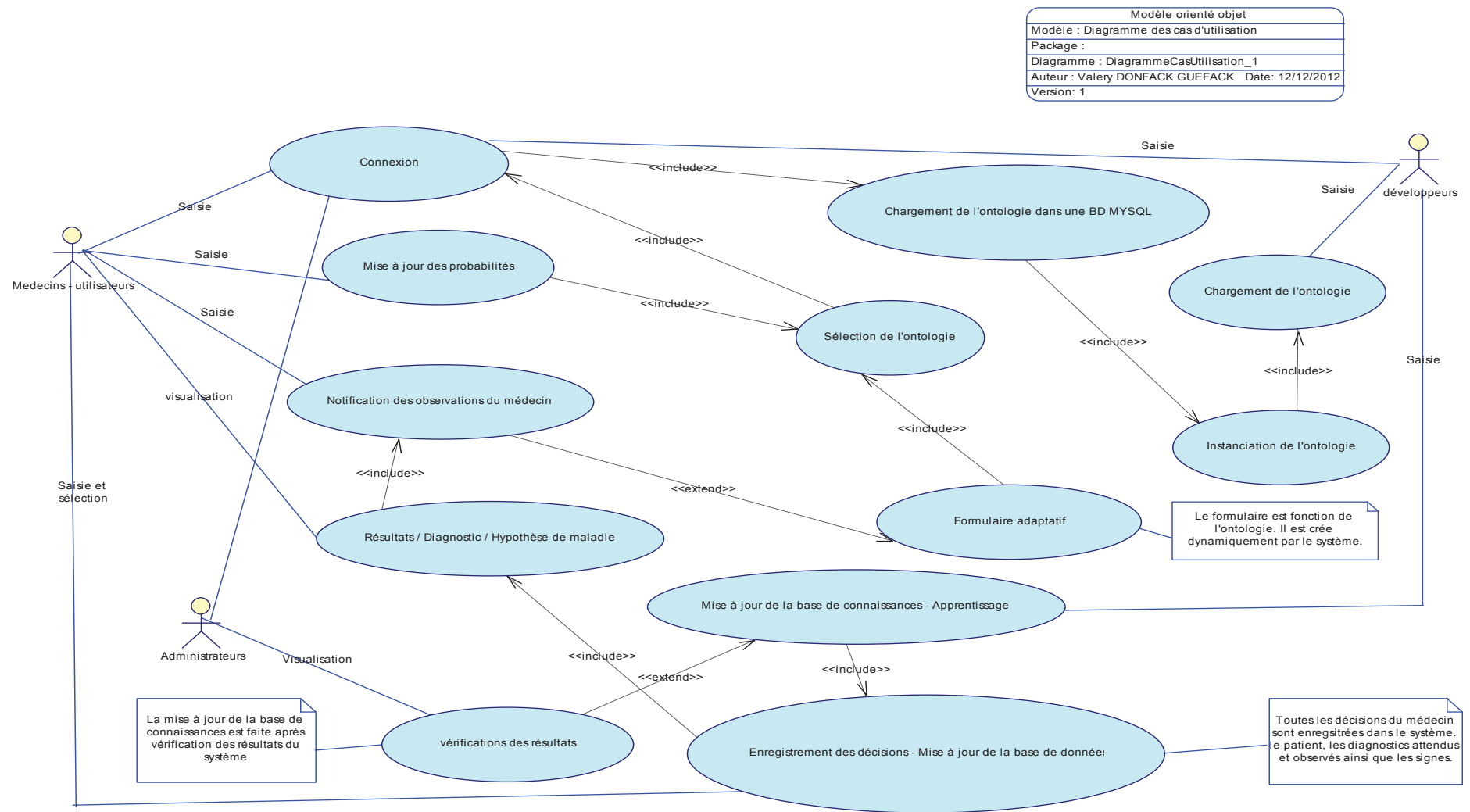


Figure 26: Diagramme des cas d'utilisation du système ODES. Les relations d'utilisation « include » stipulent que le cas d'utilisation source contient aussi le comportement décrit dans le cas d'utilisation destination. En d'autre terme pour réaliser l'objectif « sélection de l'ontologie », on utilise les objectifs « connexion ». Les relations d'extension « extend » stipulent que le cas d'utilisation source étend (précise) les objectifs ou le comportement du cas d'utilisation destination. En d'autres termes, l'objectif « Mise à jour de la base de connaissances - Apprentissage » est réalisé après « vérification des résultats ».

#### 4.2.9.2 Fonctionnement

Le système est composé de deux grandes parties interdépendantes. La première partie est celle de la génération des ensembles (cf. 4.2.7) à partir de l'ontologie biomédicale et de leur stockage dans une base de données relationnelle :

- (1) le chargement de l'ontologie dans le module « Ontology Loader » qui va créer les tuples (enregistrements) correspondants aux ensembles (cf. 4.2.7). (1') Les tuples sont enregistrés dans la base de données relationnelle ou base de connaissances relationnelle (RKB).
- (2) le module « Interpreter for interface » extrait les signes et leur catégorie ou famille et génère le formulaire adaptatif. Ce module teste, si le signe est dichotomique ou bivalent alors il dessine des cases à cocher, sinon, si le signe est pluri-catégoriel ou non bivalent alors il dessine une zone de texte ou une liste déroulante.

La deuxième partie est celle de l'exploitation de cette base :

- (3) Saisie des signes et symptômes (paramètres) d'un patient par un utilisateur du système,
- (4) Prise en charge des paramètres saisies par le module « Interpreter ». Il crée les listes correspondant au contexte (saisie des observations faites sur un patient ou enregistrement des décisions des utilisateurs) :
  - a. Une liste des signes présents,
  - b. Une liste des signes absents,
  - c. Une liste des décisions de l'utilisateur.
- (5) Élaboration de la requête (extraction ou stockage) à partir de la liste des signes ou des décisions.
- (6) Interrogation de la RKB.
- (7) Réponse de la RKB.
- (8) Envoi des résultats à l'« Interpreter ». Toutes les hypothèses de maladies, les signes associés et les relations (hypothèses de maladies/signes) sont triées soit via les ratios ou via les probabilités post-tests. Les signes associés aux hypothèses de maladies et les relations (hypothèses de maladies/signes) sont aussi envoyés à l'« Interpreter ».
- (9) Les signes présents, les signes absents et les signes non observés qui caractérisent chaque hypothèse de maladie proposée sont affichés au niveau de l'interface (Decision support system interface).
- (10) Sélection du ou des diagnostic(s) positif(s), création de la liste des décisions de l'utilisateur (4).
- (11) Stockage de la liste des décisions de l'utilisateur dans la base de cas « Case Base ».
- (12) Interrogation de la base de cas par le module « DSS Learning unit ».
- (13) Réponse de la base de cas, envoi des nouveaux cas.
- (14) Mise à jour<sup>22</sup> de la RKB, cette mise à jour doit être supervisée (non automatique).

---

<sup>22</sup> Concerne les prévalences, les sensibilités et les spécificités.

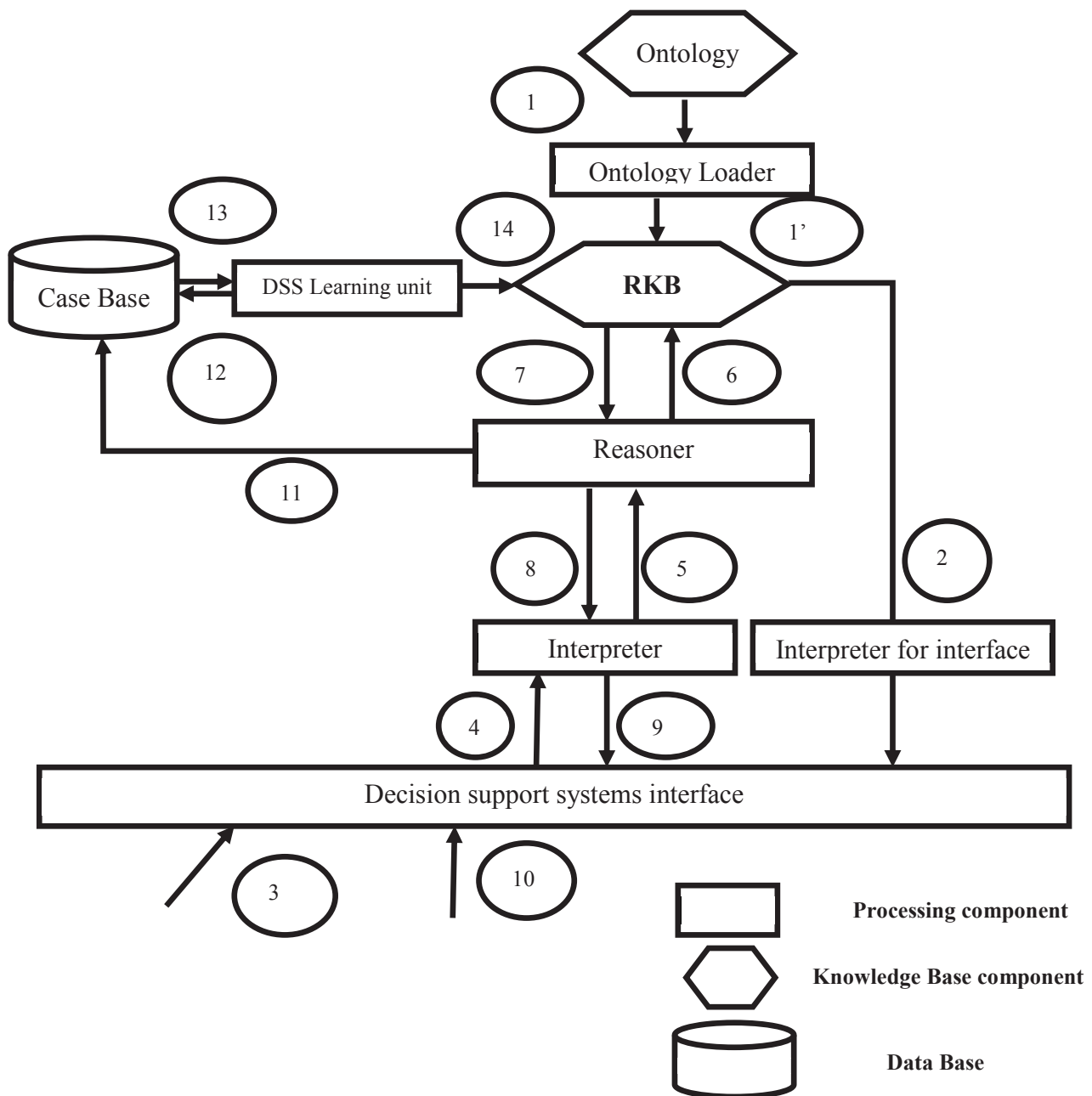


Figure 27: Vue du schéma général du système d'aide au diagnostic médical fondé sur une ontologie biomédicale. Sur ce schéma nous avons toutes les composantes du système, partant de la génération des tuples nécessaire pour l'aide au diagnostic à l'étape de mise à jour de ces tuples. Les étapes de construction de l'ontologie biomédicale ne sont pas mentionnées car elles sont faites en amont de la génération des tuples et de l'exploitation dudit système (Jovic et al., 2007).

#### 4.2.9.3 Le modèle conceptuel des données

Le modèle conceptuel des données (MCD) a pour but de décrire de façon formelle les données qui seront utilisées par le système d'information. Il s'agit donc d'une représentation des données, facilement compréhensible, permettant de décrire le système d'information (SI) à l'aide d'entités (Matheron, 2005). Il décrit la sémantique, c'est-à-dire le sens attaché à ces données et à leurs rapports et non à l'utilisation qui peut en être faite. Un MCD est constitué d'entités types, de relations et de propriétés. Une entité type est la représentation dans le système d'information d'un objet matériel ou immatériel de l'univers extérieur. C'est plus

exactement un ensemble d'objets ayant les mêmes caractéristiques. Dans une entité type, on met les informations nécessaires et suffisantes pour caractériser cette entité type. Une relation est la prise en charge par le SI du fait qu'il existe une association entre des objets de l'univers extérieur. C'est un lien entre deux entités types (ou plus). On doit lui donner un nom, souvent un verbe, qui caractérise le type de relation entre les entités types. Une propriété est une rubrique attribut d'une entité type ou d'une relation.

#### 4.2.9.4 Le modèle conceptuel de la base de connaissances

Les entités types décrivant la base de connaissances sont les suivantes :

- ❖ E\_ONTOLOGY : Les ontologies gérées par le système ODES. ODES est multi ontologique, il donne la possibilité à partir de cette entité type d'indexer toutes les maladies rattachées à chaque ontologie. Une ontologie renferme une ou plusieurs maladies et une maladie n'est rattachée qu'à une seule ontologie (cette dernière a été définie comme telle pour simplifier le modèle, car il est possible qu'une maladie soit présente dans deux ou plusieurs ontologies différentes).
- ❖ E\_DISEASE : Les maladies gérées par le système. En effet, toutes les informations sur les maladies gérées par le système sont stockées dans la base de connaissances.
- ❖ E\_CATEGORY : Les catégories ou familles de signes. Il s'agit par exemple des signes radiologiques, biologiques, etc. Ces informations sont utiles pour le formulaire adaptatif car tous les signes ou paramètres sont groupés par famille ou catégorie.
- ❖ E\_FINDING : Les signes et observations gérés par le système.
- ❖ E\_RELATION : Les relations entre signes et maladies gérées par le système. En effet, toutes les informations sur les relations entre les signes et les maladies gérées par le système sont stockées dans la base des cas.
- ❖ E\_CONCEPT\_RESTRICTION : Les restrictions de l'ontologie qui contiennent des valeurs seuils. C'est le cas, par exemple de l'IgA\_Measurement < 3g/dL pour le cas d'une MGUS.
- ❖ E\_CONCEPT\_RESTRICTION\_1 : Les restrictions de l'ontologie. Nous avons ajouté à ce niveau les sensibilités et spécificités des signes pour les maladies.
- ❖ E\_CONTEXT : Les contextes de l'ontologie. Ces contextes sont utiles pour caractériser les valeurs des prévalences qui sont fonctions de ces dernières. Par exemple, le contexte « patients âgés de plus de 70 ans ». Le contexte « patients habitant en Bretagne ».
- ❖ E\_CONTEXT\_DISEASE : La prévalence d'une maladie selon un contexte. Par exemple, la prévalence du MGUS est fonction de l'âge des patients (Watanaboonyongcharoen et al., 2012).



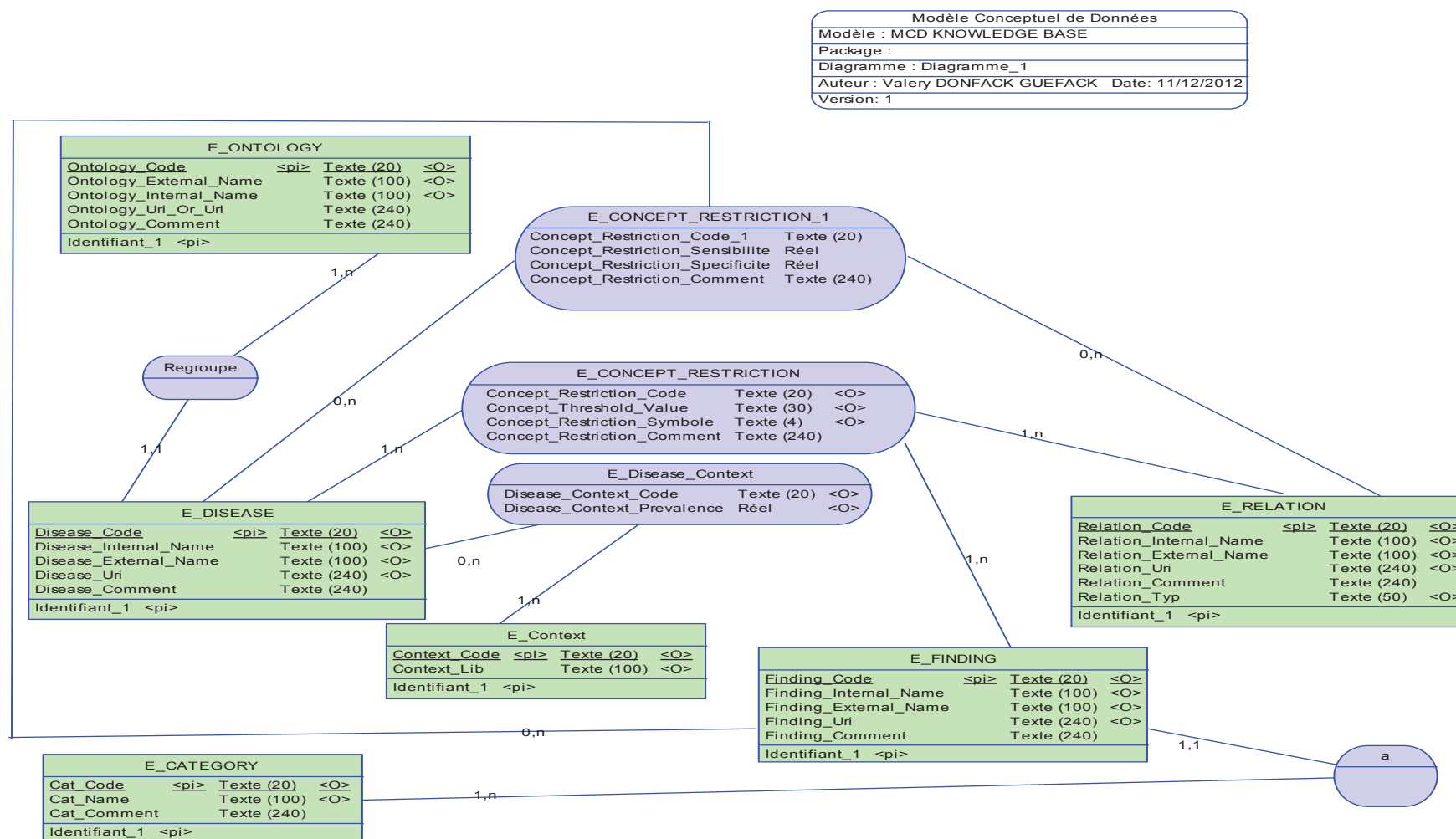


Figure 28: Modèle conceptuel de la base de connaissances du système. Cette base de connaissances est utilisée pour stocker le contenu de l'ontologie sémiologique biomédicale. C'est une sorte de copie de l'ontologie, toutes les informations nécessaires pour la description des maladies y sont. Ce choix a été fait uniquement pour des raisons de performance et de facilité d'accès pour les tests et évaluations multicentriques.

#### 4.2.9.5 Le modèle conceptuel de la base des cas

Les entités types décrivant la base des cas sont les suivantes :

- ❖ E\_GROUP : Les groupes d'utilisateur qui peuvent avoir accès au système. Par exemple, administrateur, médecin, étudiant, etc.
- ❖ E\_USER : Les utilisateurs du système. Un utilisateur ne peut être que dans un seul et unique groupe.
- ❖ E\_PATIENT : Les patients gérés par le système. Nous n'avons intégré que les informations de bases des patients (Figure 29).
- ❖ E\_DIAGNOSIS : Les passages ou les événements associés à un patient. Toutes les décisions prises par les médecins y sont rattachées.
- ❖ E\_DISEASE\_DIAGNOSIS : Les diagnostics d'un patient. Toutes les décisions prises par les médecins sont stockées dans la base des cas. Et pour chaque décision on associe les signes impliqués en notifiant l'état (présent, absent) de ces derniers.
- ❖ E\_DIAG\_FINDING\_RELATION : Les signes (présents, absents) rattachés aux décisions prises par un utilisateur.

#### 4.2.9.6 Le calcul de la sémiologie quantitative

Toutes les décisions des utilisateurs (médecins, cliniciens, etc.) sont stockées dans la base de cas (Figure 29). Chaque décision ou hypothèse de maladie est enregistré avec les signes présents et absents qui y sont rattachés (signes qui l'impliquent). A partir de ces enregistrements, les effectifs (cf. 2.4) sont calculés, et à l'aide de ces effectifs les sensibilités et spécificités des signes pour les maladies sont produites, ainsi que les prévalences des maladies pour la population considérée.

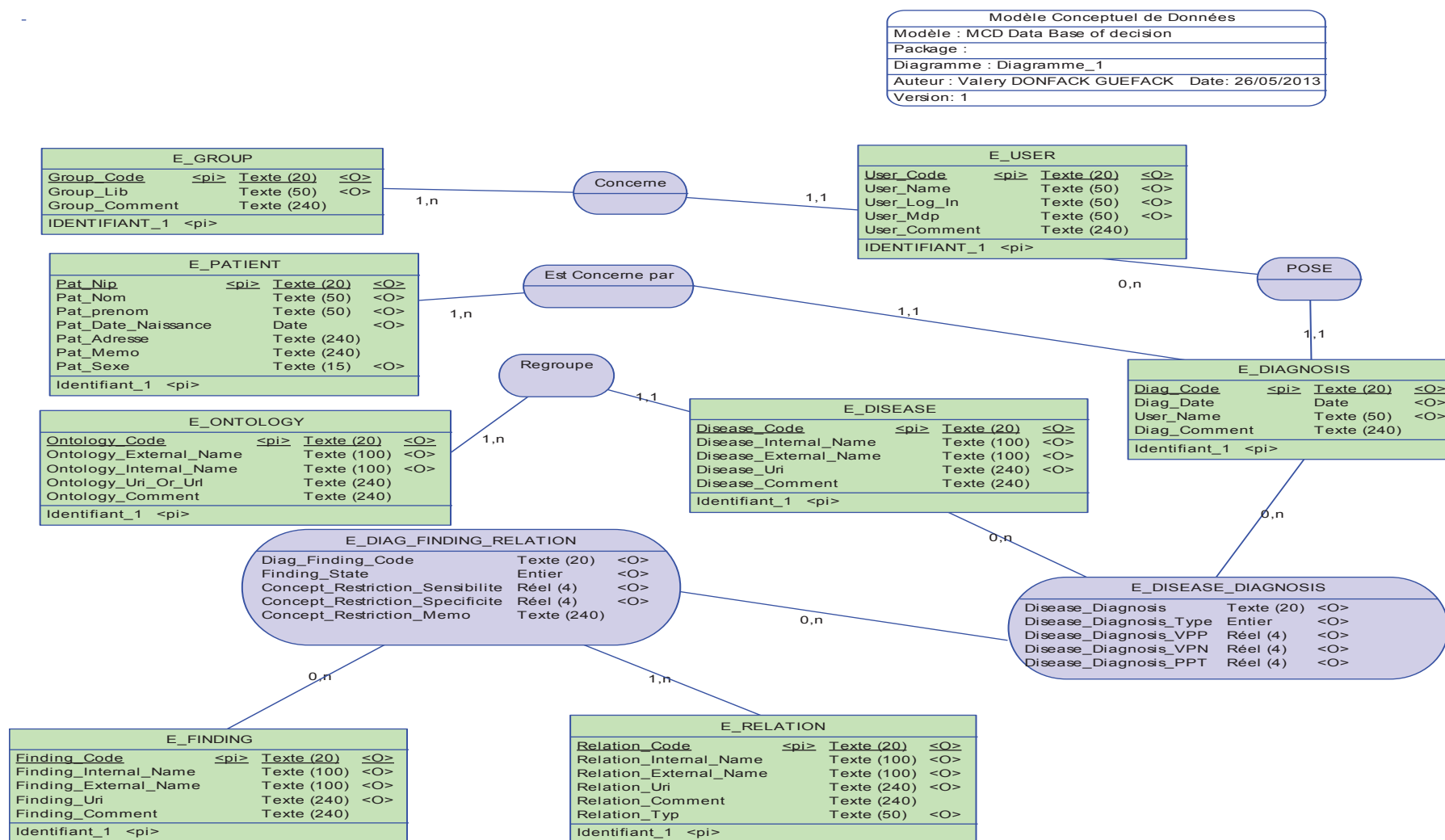


Figure 29: Modèle conceptuel de la base des cas du système. Cette base des cas est utilisée pour stocker les informations sur les décisions prises par les médecins. Ces données sont ensuite utilisées pour l'évaluation des performances du système et aussi pour affiner le modèle de connaissance via le processus d'apprentissage. Pour garantir l'intégrité et la cohérence de la base des cas, des entités types représentant les maladies, les signes et les relations ont été intégrées pour stocker le contenu de l'ontologie sémiologique biomédicale et ainsi conserver à ce niveau une partie de la sémantique du domaine.

#### 4.2.9.7 Dictionnaire des données

Un dictionnaire des données est une collection de métadonnées nécessaires à la conception d'une base de données relationnelle (Matheron, 2005).

**Tableau 10: Le dictionnaire de données.**

| Nom                    | Signification   | Type        | Nature                |
|------------------------|---|-------------|-----------------------|
| <u>Group_Code</u>      | Code du groupe d'utilisateur  | Texte (20)  | Obligatoire et unique |
| Group_Lib              | Libellé du groupe   | Texte(50)   | Obligatoire           |
| Group_Comment          | Commentaire ou explication  | Texte(240)  | facultatif            |
| <u>User_Code</u>       | Code de l'utilisateur   | Texte(20)   | Obligatoire           |
| User_Name              | Nom de l'utilisateur  | Texte(50)   | Obligatoire           |
| User_Log_In            | Nom de connexion de l'utilisateur   | Texte(50)   | Obligatoire et unique |
| User_Mdp               | Mot de passe de l'utilisateur   | Texte(50)   | Obligatoire           |
| User_Comment           | Commentaire sur le poste ou la spécialité de l'utilisateur, sachant que ce dernier peut être un médecin, etc. | Texte(240)  | Facultatif            |
| <u>Ontology_Code</u>   | Code de l'ontologie   | Texte (20)  | Obligatoire           |
| Ontology_External_Name | Nom externe de l'ontologie, celui qui sera affiché sur le formulaire  | Texte (100) | Obligatoire et unique |
| Ontology_Internal_Name | Nom interne manipulé par le système   | Texte (100) | Obligatoire et unique |
| Ontology_Uri_Or_Url    | Uri ou Url de l'ontologie   | Texte(240)  | Facultatif            |
| Ontology_Comment       | Commentaire ou explication  | Texte(240)  | Facultatif            |
| <u>Relation_Code</u>   | Code de la relation   | Texte(20)   | Obligatoire et unique |
| Relation_External_Name | Nom externe de la relation, celui qui sera affiché sur le formulaire  | Texte(100)  | Obligatoire et unique |
| Relation_Internal_Name | Nom interne de la relation, celui qui est manipulé par le système   | Texte(100)  | Obligatoire et unique |
| Relation_Uri           | Uri de la relation  |             | facultatif            |
| Relation_Comment       | Commentaire sur la description et l'objet de la relation  | Texte(240)  | facultatif            |
| Relation_Typ           | Concerne le type de relation, si c'est un objectproperty ou un datatypeproperty                               | Texte(50)   | Obligatoire           |
| <u>Disease_Code</u>    | Code de la maladie  | Texte(20)   | Obligatoire et unique |
| Disease_External_Name  | Nom externe de la maladie, celui qui sera affiché sur le formulaire   | Texte(100)  | Obligatoire et unique |
| Disease_Internal_Name  | Nom interne de la maladie, celui qui est manipulé par le système  | Texte(100)  | Obligatoire et unique |
| Disease_Uri            | Uri de la maladie   | Texte(240)  | facultatif            |
| Disease_Comment        | Commentaire supplémentaire sur la description de la maladie   | Texte(240)  | facultatif            |
| Finding_Code           | Code du signe, symptôme   | Texte(20)   | Obligatoire et unique |
| Finding_External_Name  | Nom externe du signe, celui qui sera affiché sur le formulaire  | Texte(100)  | Obligatoire et unique |
| Finding_Internal_Name  | Nom interne du signe, celui qui   | Texte(100)  | Obligatoire et        |

|                                   |   |              |  |
|-----------------------------------|---|--------------|--|
|                                   | est manipulé par le système   |              | unique                                   |
| Finding_Uri                       | Uri du signe  | Texte(240)   | facultatif                               |
| Finding_Comment                   | Commentaire supplémentaire sur la description du signe  | Texte(240)   | facultatif                               |
| Pat_Nip                           | Numéro d'identification du patient.   | Texte(20)    | Obligatoire et unique                    |
| Pat_Nom                           | Nom du patient  | Texte(50)    | Obligatoire                              |
| Pat_Prenom                        | Prénom du patient   | Texte(50)    | Obligatoire                              |
| Pat_Date_Naissance                | Date de naissance du patient  | Date         | Obligatoire                              |
| Pat_Adresse                       | Adresse du patient  | Texte(240)   | Facultatif                               |
| Pat_Memo                          | Commentaire supplémentaire sur le patient   | Texte(240)   | Facultatif                               |
| Pat_Sexe                          | Sexe du patient   | Texte(15)    | Obligatoire                              |
| Diag_Code                         | Numéro du passage ou de l'événement.  | Texte(20)    | Obligatoire et unique                    |
| Diag_Date                         | Date d'enregistrement de l'événement.   | Date         | Obligatoire                              |
| Diag_Comment                      | Commentaire supplémentaire sur l'événement.   | Texte(240)   | Facultatif                               |
| Disease_Diagnosis                 | C'est un code permettant d'identifier l'événement et le diagnostic associé.   |              | Obligatoire et unique                    |
| Disease_Diagnosis_Type            | Type de diagnostic. Si c'est le diagnostic retenu par le médecin ou un diagnostic attendu, c'est-à-dire une des propositions du système.  | Entier (0,1) | Obligatoire                              |
| Disease_Diagnosis_VPP             | Valeur prédictive positive du diagnostic.   | Réel         | Obligatoire                              |
| Disease_Diagnosis_VPN             | Valeur prédictive négative du diagnostic. Par défaut, nous mettons zéro.  | Réel         | Obligatoire                              |
| Disease_Diagnosis_PPT             | Valeur prédictive du diagnostic. Cette valeur tient compte des signes présents et absents à la fois. Dans le cas où on ne prendrait en compte que les ratios, c'est le ratio qui y serait stocké. | Réel         | Obligatoire                              |
| Disease_Diagnosis_Memo            | Commentaire relatif au choix d'un médecin pour une hypothèse de maladie donnée.   | Texte(240)   | Facultatif                               |
| Diag_Finding_Code                 | C'est un code utile pour le signe. C'est le signe de la maladie d'un patient donné.   | Entier       | Obligatoire et unique                    |
| Concept_restriction_Finding_State | Etat du signe, présent ou absent  | Entier       | Obligatoire                              |
| Concept_restriction_Memo          | Commentaire relatif au signe impliquant ou excluant une hypothèse de maladie.   | Texte(240)   | Obligatoire, mais inséré par le système. |
| Concept_Restriction_Code_1        | Code d'une restriction.   | Texte (20)   | Obligatoire et unique                    |
| Concept_restriction_Sensibilite   | C'est la valeur de la sensibilité d'un signe pour une maladie.  | Réel         | Obligatoire                              |
| Concept_Restriction_Specificite   | C'est la valeur de la spécificité d'un signe pour une maladie.  | Réel         | Obligatoire                              |
| Concept_Restriction_Comment       | Commentaires supplémentaires sur la restriction.  | Texte(240)   | Facultatif                               |
| Concept_Restriction_Code          | Code d'une restriction.   |              | Obligatoire et unique                    |

|                                     |   |            |                       |
|-------------------------------------|---|------------|-----------------------|
| Concept_Restriction_Threshold_Value | Valeur du signe dans le cadre de l'observation d'un signe multi-value ou pluri catégoriel, comme le taux d'hémoglobine ou le stade d'un myélome multiple. | Texte(240) | Obligatoire           |
| Concept_Restriction_Symbole         | Symbole (=, >, <, ≥, ≤, ≠)  | Texte(4)   | Obligatoire           |
| Disease_Context_Code                | Identifiant de la prévalence d'une maladie selon un contexte.   | Texte(20)  | Obligatoire et unique |
| Disease_Context_Prevalence          | Valeur de la prévalence d'une maladie selon le contexte. Par défaut, on met la valeur zéro.   | Réel       | Obligatoire           |
| Context_Code                        | Code du contexte.   | Texte(20)  | Obligatoire et unique |
| Context_Lib                         | Nom ou le libellé du contexte   | Texte(100) | Obligatoire et unique |
| <u>Cat_Code</u>                     | Code de la catégorie  | Texte(20)  | Obligatoire et unique |
| Cat_Name                            | Nom de la catégorie   | Texte(100) | Obligatoire et unique |
| Cat_Comment                         | Commentaires supplémentaires sur la description ou l'objet de la catégorie  | Texte(240) | Facultatif            |

#### 4.2.10 Évaluation

L'évaluation est une démarche qui vise à mesurer, quantifier (méthodes statistiques) et caractériser une situation, une entité, un résultat ou une performance de nature complexe et donc *a priori* difficilement mesurable (Nakache & Metais, 2005). Elle consiste à mesurer la différence entre un résultat attendu et un résultat obtenu afin de définir la performance d'un système. La performance est la mesure de la capacité d'un système à satisfaire un utilisateur en termes de pertinence des résultats renvoyés (dans notre cas des diagnostics).

##### 4.2.10.1 Données médicales

Les dossiers patients ont été recueillis dans deux services du CHU de Rennes et un service de l'hôpital sud de Rennes. Au CHU de Rennes nous avons recueilli 202 dossiers patients au service des urgences odontologiques et 20 dossiers patients au service de médecine du sport (pathologies traumatiques du genou). A l'hôpital sud de Rennes nous avons recueilli 63 dossiers patients au service de médecine interne.

##### 4.2.10.2 Méthodes d'évaluation

Les résultats renvoyés par le système sont contenus dans deux listes, une liste des diagnostics possibles et une liste des diagnostics exclus. Sa performance est alors caractérisée par les taux de diagnostics corrects ou précision (RCD), de diagnostics incorrects (ER) et de diagnostics non discriminants (RDnD).

Un diagnostic correct (DC) est, si les probabilités sont absentes, un diagnostic présent dans la liste des diagnostics possibles renvoyés par le système quelle que soit sa position dans cette liste. Et, si nous disposons des probabilités, un diagnostic est considéré comme étant correct s'il est le premier dans la liste des diagnostics possibles.

Un diagnostic incorrect (DI) quant à lui est un diagnostic non renvoyé par le système ou un diagnostic renvoyé, mais dans la mauvaise liste (par exemple, liste des diagnostics exclus).

Un diagnostic non discriminant (DnD) est un diagnostic possible équiprobable avec le ou les diagnostics retenus, cette catégorie de diagnostics concerne les cas où nous avons les probabilités.

Un diagnostic retenu est celui retenu par le médecin et donc celui dont souffre potentiellement le patient.

Les résultats obtenus par le système ont été comparés à ceux obtenus par les médecins, cliniciens séniors, et dans certains cas à ceux de l'IRM (Imagerie par Résonance Magnétique) afin de déterminer les taux de classifications correctes, incorrectes et non discriminantes pour chaque série de dossiers.

Ces taux sont calculés en divisant le nombre de cas de chaque catégorie (DC, DI et DnD) renvoyé par le système par le nombre total des cas testés (n) (De Dombal, Leaper, Horrocks, Staniland, & McCann, 1974; Essex & Healy, 1994; J. Horrocks et al., 1972; Lemaire et al., 1999; Scarlett, Cooke, Clarke, Bates, & Chan May, 1986). Pour les résultats de l'évaluation (cf. 5.5).

$$\mathbf{RCD} = \frac{\mathbf{DC}}{\mathbf{n}} \quad (\text{Eq. 1})$$

$$\mathbf{ER} = \frac{\mathbf{DI}}{\mathbf{n}} \quad (\text{Eq. 2})$$

$$\mathbf{RDnD} = \mathbf{DnD}/\mathbf{n} \quad (\text{Eq. 3})$$

### 4.3 Conclusion

Dans ce chapitre, nous avons présenté le matériel et méthode de ce travail. Nous avons d'abord présenté le corpus utilisé pour la mise en œuvre des méthodes de modélisation permettant d'obtenir une ontologie sémiologique biomédicale (ontologie biomédicale adaptée au processus de diagnostic médical). Ensuite, nous avons présenté notre approche pour enrichir ou compléter les connaissances dans nos modèles ontologiques. Puis, nous avons présenté l'architecture du système d'aide au diagnostic à base d'ontologie (ODES) et enfin nous avons présenté la méthode d'évaluation du système.





## Chapitre 5

### Résultats

#### 5.1 Introduction

Le modèle ontologique décrivant les maladies et le système d'aide au diagnostic médical sont présentés. Ces résultats sont séparés en trois parties : la première concerne le modèle ontologique, la deuxième concerne le système ODES et la troisième traite des résultats de l'évaluation. Les illustrations sont faites en s'appuyant sur les exemples de la gammopathie monoclonale à signification indéterminée (MGUS) et du myélome multiple (MM) tirés des articles (Kyle et al., 2011; Rajkumar, 2011). Toutefois, les résultats de l'évaluation tiennent compte des trois domaines d'application que nous avons présentés précédemment (cf. 4.1.1.2, 4.1.1.3 et 4.1.1.4).

Concernant le modèle ontologique, nous présentons la représentation : (a) dichotomique du signe, (b) multivaluée du signe, (c) de la sémiologie quantitative, (d) des signes absents ou négatifs, (e) multi-niveaux des maladies et, (f) des connaissances de raisonnement ou de contrôle. Une base de règles d'ordre 0 est produite à partir du contenu de l'ontologie instanciée et le moteur d'inférences JESS.

Concernant le système, nous présentons : (a) l'élaboration de la base de connaissances dite base de connaissances relationnelle, (b) le moteur d'inférences en mettant une emphase sur les modes de raisonnement implémentés (abductif et probabiliste), (c) le module de génération de formulaire adaptatif.

Pour ce qui est de l'évaluation, les résultats pour chaque domaine d'application sont présentés (cf. 5.5).

#### 5.2 Le modèle de connaissances

##### 5.2.1 La représentation dichotomique du signe

Le signe est discrétisé sous forme de deux sous-classes représentant les deux modalités (Figure 30).

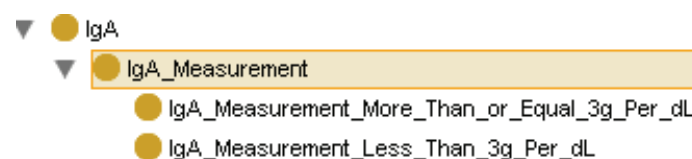


Figure 30: Dans le cadre de la représentation des critères diagnostiques d'une MGUS, il est nécessaire de distinguer si la mesure de l'IgA est inférieure à 3g/dL ou non. Dans ce cas, la mesure de l'IgA peut être représentée par deux sous-classes représentant pour l'une une mesure d'IgA < 3g/dL et pour l'autre une mesure d'IgA ≥ 3g/dL.

Une des deux modalités peut ensuite être utilisée comme critère diagnostique dans la définition formelle de la maladie (Figure 31).

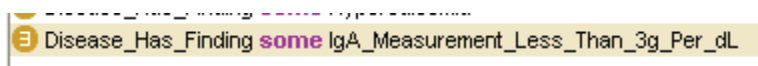


Figure 31: Définition formelle partielle d'une MGUS.

### 5.2.2 La représentation multivaluée du signe

Il s'agit du cas où un signe peut prendre une multitude de valeurs. Cette connaissance met en jeu une relation d'arité supérieure ou égale à 3 (un signe dans une maladie donnée prend une valeur avec une unité ou pas). OWL ne permettant de formaliser que des relations unaires ou binaires, nous proposons de réifier la relation maladie-signes à laquelle nous attribuons la valeur et l'unité du signe dans cette maladie (Figure 21). La représentation de ce type de connaissance nécessite donc : (a) la réification de la relation maladie-signes, (b) la création des relations de type *ObjectProperty* (Figure 32) permettant de relier la relation précédemment réifiée à sa maladie (*Has\_Disease*) et à son signe (*Has\_Finding*) (Figure 33).

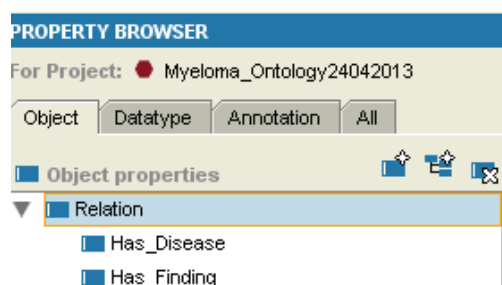


Figure 32: Création de nouvelles relations de type « *ObjectProperty* ».

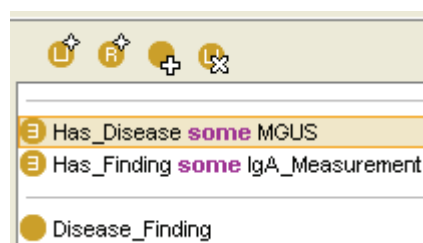


Figure 33: Définition partielle de la classe MGUS\_IgA\_Measurement.

(c) La création des relations de type « *DatatypeProperty* » (Figure 34) permettant de relier la relation précédemment réifiée à la valeur et à l'unité de la mesure du signe (Figure 35).

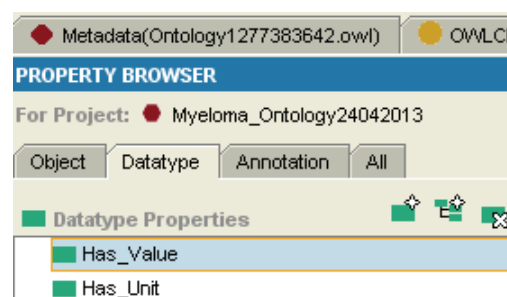


Figure 34: Création de nouvelles relations de type « *DatatypeProperty* ».

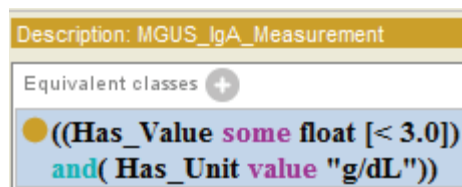


Figure 35: Définition formelle partielle de la classe MGUS\_IgA\_Measurement.

### 5.2.3 La représentation de la probabilité d'un signe dans une maladie

Il s'agit du cas où la survenue d'un signe  $s$  pour une maladie  $m$  a une probabilité  $p$ . Cette connaissance met en jeu une relation d'arité supérieure à 2 (un signe est présent dans une maladie avec une probabilité donnée.). OWL ne permettant de formaliser que des relations unaires ou binaires, nous proposons de réifier la relation maladie-signes à laquelle nous attribuons des probabilités, par exemple la sensibilité et la spécificité (cf. 2.4).

Cette représentation nécessite donc : (a) la réification de la relation maladie-signes, (b) la création des relations de type `ObjectProperty` (Figure 32) permettant de relier la relation précédemment réifiée à sa maladie (`Has_Disease`) et à son signe (`Has_Finding`) (Figure 36).

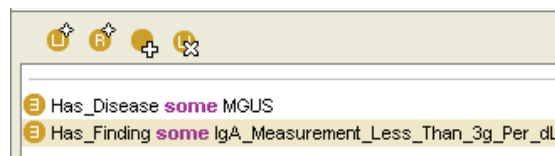


Figure 36: Définition formelle partielle de la classe MGUS\_IgA\_Measurement\_Less\_Than\_3g\_Per\_dL.

(c) La création des relations de type `DatatypeProperty` (Figure 37) permettant de relier la relation précédemment réifiée (Figure 23) aux valeurs de la sensibilité et de la spécificité du signe pour la maladie (Figure 38).

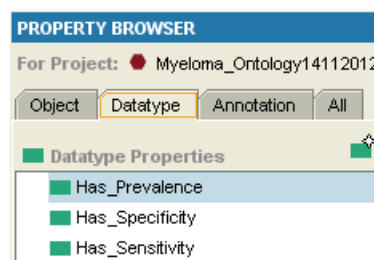


Figure 37: Création de nouvelles relations de type `DatatypeProperty` pour la représentation des prévalences des maladies pour un contexte et, des sensibilités et des spécificités des signes pour les maladies.

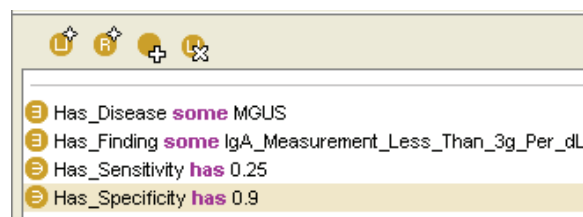


Figure 38: Définition formelle partielle de la classe MGUS\_IgA\_Measurement\_Less\_Than\_3g\_Per\_dL.

### 5.2.4 La représentation de la prévalence d'une maladie

Il s'agit de la probabilité  $p$  d'avoir une maladie  $m$  dans une population donnée. La population dans notre cas est matérialisée par la notion de contexte. Le contexte est un concept abstrait qui peut avoir pour instance : les patients âgés de plus de 70 ans, les patients vivant en Bretagne, etc. Cette connaissance met en jeu une relation d'arité supérieure à 2 (Figure 24). Nous proposons de réifier la relation maladie-contexte à laquelle nous attribuons une prévalence (cf. 2.4).

Cette représentation nécessite donc : (a) la réification de la relation maladie-contexte (Figure 24), (b) la création des relations de type `ObjectProperty` (Figure 39) permettant de relier la relation précédemment réifiée à son contexte (`Has_Context`) et à sa maladie (`Has_Disease`) (Figure 40).

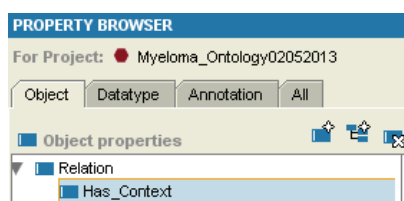


Figure 39: Création d'une nouvelle relation de type « `ObjectProperty` ».

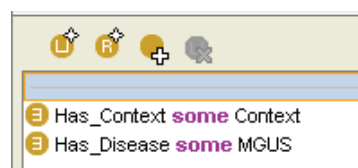


Figure 40: Définition formelle partielle de la classe `MGUS_Context`.

(c) La création d'une relation de type `DatatypeProperty` (Figure 37) permettant de relier la relation précédemment réifiée (Figure 24) à la valeur de la prévalence (Figure 41).

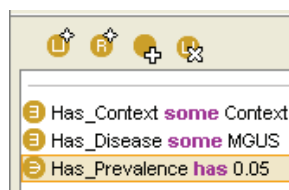


Figure 41: Définition formelle de la classe `MGUS_Context`.

#### 5.2.4.1 La représentation multi-niveaux des maladies

Il s'agit des cas où certaines maladies sont des signes pour d'autres maladies (cf. 4.2.1.4). Pour les illustrations nous prendrons le cas du myélome multiple (MM) (Figure 25).

Cette représentation nécessite : (a) l'expression de la restriction entre l'hypercalcémie et la mesure du calcium sérique (Figure 42), et (b) l'expression de la restriction entre le MM et l'hypercalcémie (Figure 43). Les relations utilisées dans ces définitions sont dans le tableau 3.

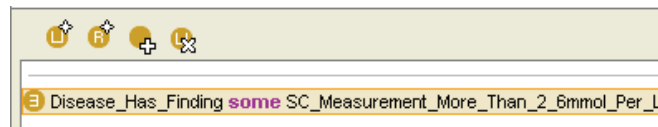


Figure 42: Définition formelle partielle de l'hypercalcémie.

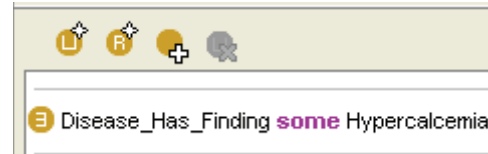


Figure 43: Définition formelle partielle du myélome multiple.

#### 5.2.4.2 La représentation des signes absents

Notre approche de représentation des signes absents ou négatifs (cf. 2.1.2 et 4.2.2) repose sur l'utilisation des relations (Figure 44). Elle s'inspire des travaux effectués sur l'ontologie du NCIT et de Werner Ceuster.

La présence de lésions osseuses « exclut » l'éventualité d'une MGUS (Rajkumar, 2011), pour l'exprimer nous utilisons une relation de type « Excludes » (Tableau 3).



Figure 44: Définition formelle partielle d'une MGUS avec un signe absent ou négatif.

#### 5.2.5 La représentation des connaissances de contrôle ou de raisonnement

Il s'agit des connaissances utilisées par le moteur d'inférences. Elles guident le moteur d'inférences en définissant quand et comment utiliser une connaissance ou un fragment de connaissance (cf. 3.2.3) (par exemple, en cas de présence ou d'absence d'un signe chez un patient donné, utiliser telles connaissances, etc.). Notre proposition consiste à les exprimer avec des règles SWRL (cf. 4.2.4). Deux grandes catégories de règles sont mises en œuvre : (a) des règles d'arité 2 via l'onglet SWRLTab (cf. 3.3.2.6) et, (b) des règles ayant des arités supérieures à 2 ou d'arité n via l'onglet SWRL-IQ (cf. 4.1.3.3). La mise en œuvre de ces règles nécessite donc : (a) la création de nouvelles relations de type « ObjectProperty » (Figure 45),

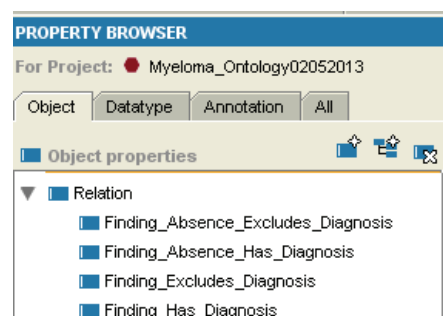


Figure 45: Création de nouvelles relations de type « ObjectProperty » pour la mise en œuvre des règles SWRL.

(b) la mise en œuvre des règles (Figure 46) et (Figure 47).

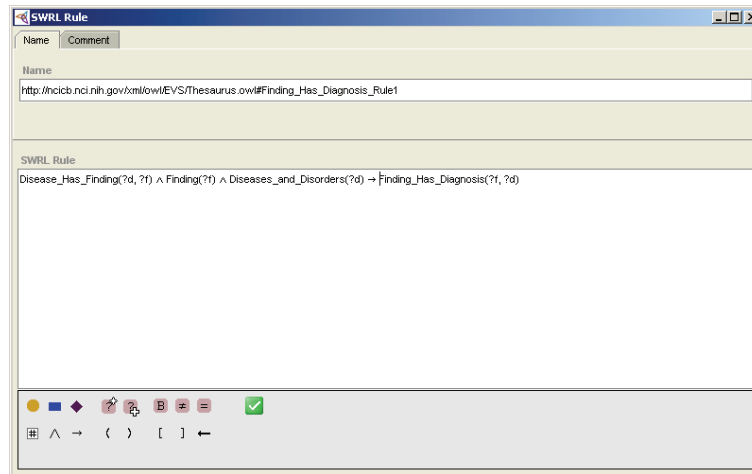


Figure 46: Expression d'une règle d'arité 2 pour des signes dichotomiques. Cet exemple concerne la règle 1.

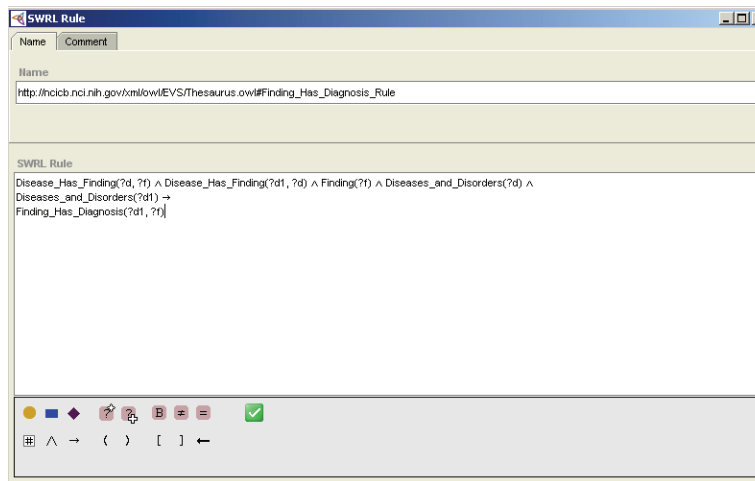


Figure 47: Règle d'arité 2 conçu via un chaînage des relations. Il permet aussi de raisonner lorsqu'on a des cas de définitions multi-niveaux des maladies. Cet exemple concerne la règle 9.

**Has\_Finding(?df,?f)^Disease\_Finding(?df)^Finding(?f)^Has\_Sensitivity(?df,?se)^Has\_Specificity(?df,?sp)^Has\_Disease(?df,?d)^Has\_Prevalence(?dc,?prev1)^Has\_Context(?dc,?c)^Context(?c)^Disease\_Context(?dc)^Has\_Disease(?dc,?d)^Diseases\_And\_Disorders(?d)^swrlb:subtract(?diff,100,?sp)^swrlb:divide(?Lrplus,se,?diff)->Finding\_Diagnosis\_Lrplus(?f,?d,?c,?prev1,?se,?sp,?Lrplus).**

**Has\_Finding(?df,?f)^Disease\_Finding(?df)^Finding(?f)^Has\_Sensitivity(?df,?se)^Has\_Specificity(?df,?sp)^Has\_Disease(?df,?d)^Has\_Prevalence(?dc,?prev1)^Has\_Context(?dc,?c)^Context(?c)^Disease\_Context(?dc)^Has\_Disease(?dc,?d)^Diseases\_And\_Disorders(?d)^swrlb:subtract(?diff,100,?se)^swrlb:divide(?Lrminus,?diff,?sp)->Finding\_Diagnosis\_Lrminus(?f,?d,?c,?prev1,?se,?sp,?Lrminus).|**

Figure 48: Expression de deux règles d'arité n. Ces règles permettent de calculer les rapports de vraisemblance (positif et négatif) et de sélectionner les autres données nécessaires pour le moteur d'inférences probabiliste. Ce type de règles est exprimé via SWRL-IQ. Cet exemple concerne respectivement les règles 11 et 12.



Les règles SWRL et le formalisme d'interrogation SWRL-IQ requièrent des instances pour fonctionner, pour cela nos modèles ontologiques doivent être instanciés (cf. 4.2.5 et 5.2.6).

### 5.2.6 La création des cas prototypiques

Nous proposons d'instancier le modèle ontologique. Cette instanciation consiste à créer une instance pour chaque classe du modèle (cf. 4.2.5). Cette instanciation est une retranscription complète des définitions intensionnelles (TBox) en définitions extensionnelles (ABox). Les noms des instances ont « \_1 » comme suffixe, par exemple l'instance de la classe MGUS sera « MGUS\_1 », celle de « Bone\_Lesion » sera « Bone\_Lesion\_1 ». Nous proposons comme exemple la définition formelle partielle d'une MGUS (Figure 49).

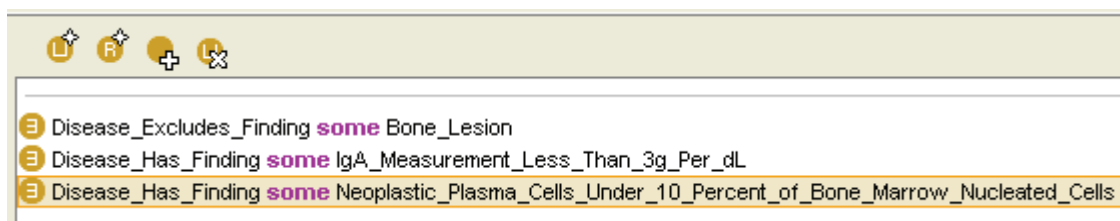


Figure 49: Définition formelle partielle d'une MGUS.

Cette instanciation nécessite donc : (a) une instance pour la classe MGUS (Figure 50) et

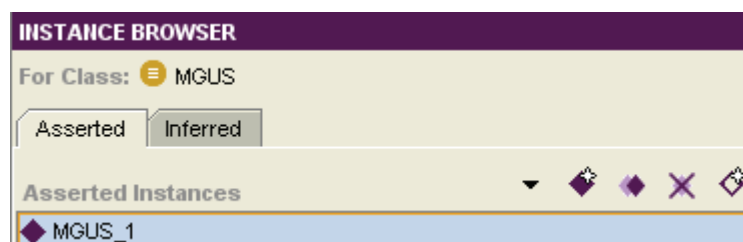


Figure 50: Instance d'une MGUS.

(b) une instance pour ses signes (Figure 51).

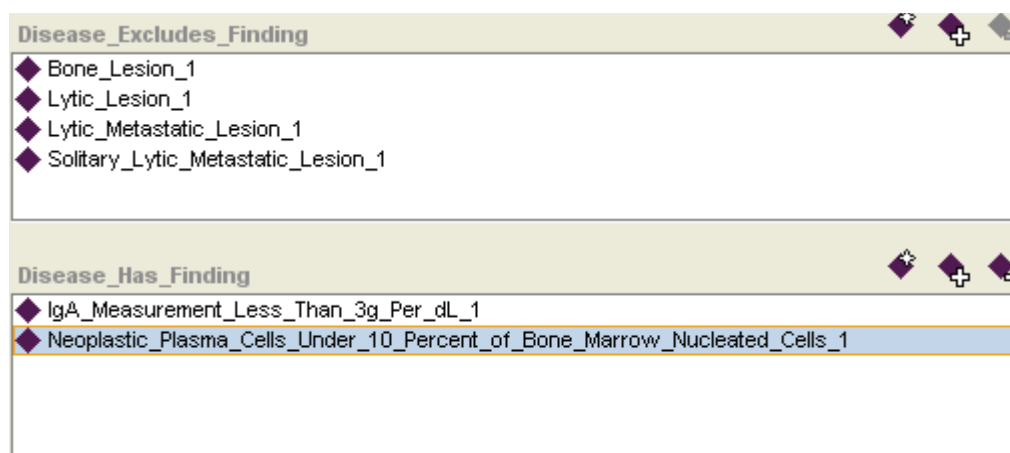


Figure 51: Instances des signes d'une MGUS.

### 5.2.6.1 La conservation de la hiérarchie des classes

Ce processus d'instanciation conserve la hiérarchie des classes. En effet, selon la relation de subsomption (Baader, 2003), si une classe D est subsumée par une classe C alors toutes les instances de la classe D sont aussi des instances de la classe C. Par exemple, la classe « Bone\_lesion » (Figure 52) a trois sous-classes « Lytic\_Lesion », « Lytic\_Metastatic\_Lesion » et « Solitary\_Lytic\_Metastatic\_Lesion ». Donc toutes les instances des classes subsumées par la classe « Bone\_Lesion » sont aussi les instances de cette dernière (Figure 51)(Figure 53).



Figure 52: Hiérarchie des types de lésion osseuse.

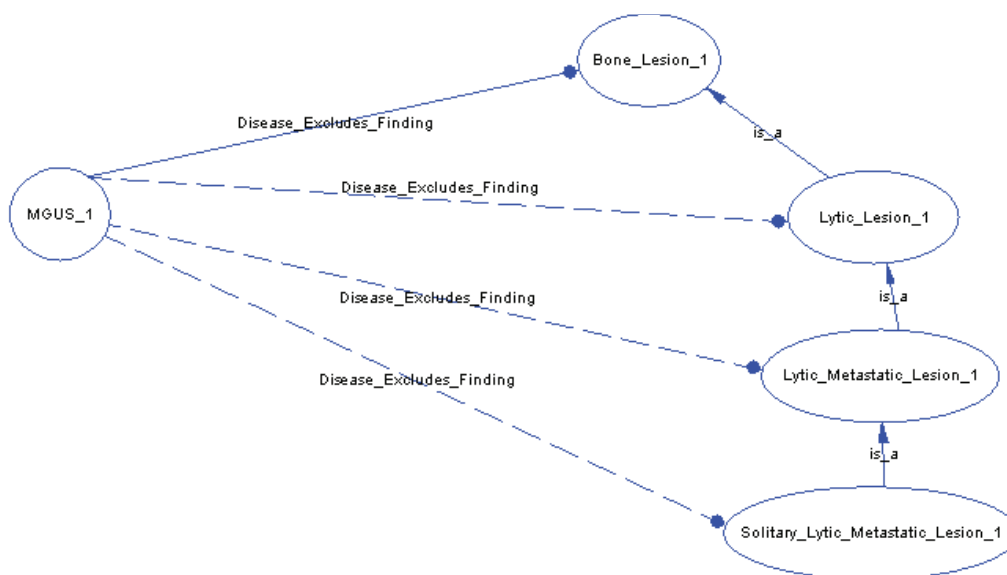


Figure 53: Illustration de la conservation de la hiérarchie des classes au niveau des instances. Toutes les instances des classes subsumées par la classe « Bone\_Lesion » sont aussi liées à l'instance « MGUS\_1 ».

### 5.2.6.2 La conservation de la profondeur des définitions

Ce processus d'instanciation conserve la profondeur des définitions des classes (Figure 25). Par exemple, le « myélome multiple » a pour signe « l'hypercalcémie », et « l'hypercalcémie » a pour signe « une mesure du calcium sérique supérieure à 2,6 mmol/L », alors par propagation transitive (Seidenberg & Rector, 2006) le « myélome multiple » a pour signe « une mesure du calcium sérique supérieure à 2.6 mmol/L » (Figure 54) (Figure 55).

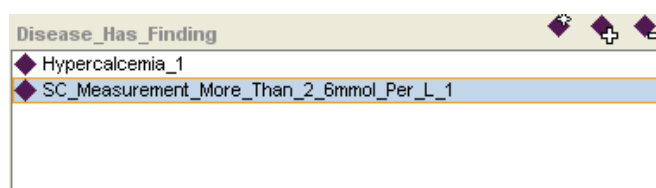


Figure 54: Instances des signes du MM. La conservation de la profondeur des définitions des classes fait en sorte que la mesure du calcium sérique est liée directement au MM.

L'instanciation dans ce contexte est une alternative au chaînage des relations mis en œuvre via les règles SWRL (cf. 3.3.2.3) (Figure 47).

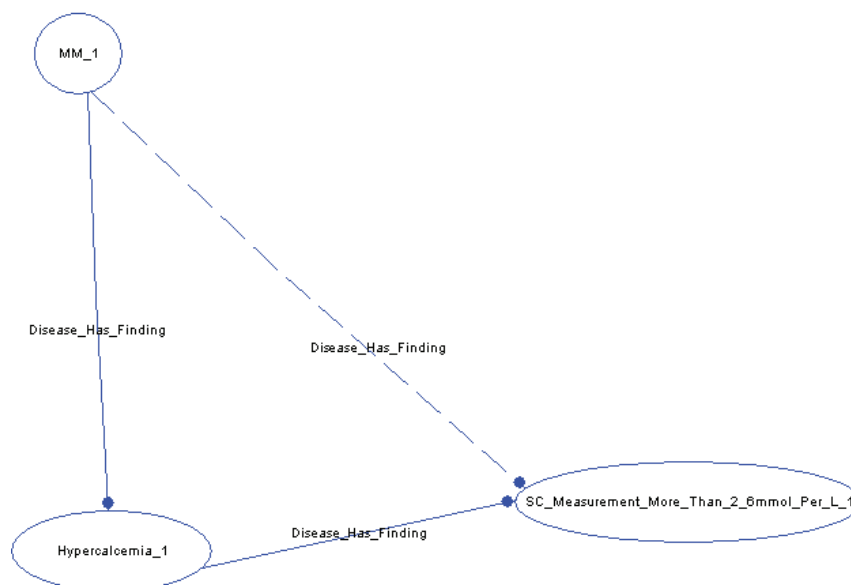


Figure 55: Illustration de la conservation de la profondeur des définitions des maladies. L'instance « MM\_1 » est liée à l'instance « Hypercalcemia\_1 » et à l'instance « SC\_Measurement\_More\_Than\_2\_6mmol\_Per\_L\_1 ».

### 5.2.6.3 La base de règles d'ordre 0

Il s'agit des règles d'ordre 0 produites sur la base des règles d'ordre 1 écrites (cf. 4.2.4), des classes et des instances de l'ontologie (cf. 5.2.6). Deux grandes catégories de règles sont produites : (a) des règles d'arité 2 (Figure 56) et (b) des règles ayant des arités supérieures à 2 ou d'arité n (Figure 57).

La production des règles d'arité 2 nécessite l'utilisation du moteur de règle JESS (cf. 4.1.3.1).

```
Finding_Has_Diagnosis [IgA_Measurement_Less_Than_3g_Per_dL_1, MGUS_1]
Finding_Has_Diagnosis [Neoplastic_Plasma_Cells_Under_10_Percent_of_Bone_Marrow_Nucleated_Cells_1, MGUS_1]
Finding_Excludes_Diagnosis [Bone_Lesion_1, MGUS_1]
Finding_Excludes_Diagnosis [Lytic_Lesion_1, MGUS_1]
Finding_Excludes_Diagnosis [Lytic_Metastatic_Lesion_1, MGUS_1]
Finding_Excludes_Diagnosis [Solitary_Lytic_Metastatic_Lesion_1, MGUS_1]
```

Figure 56: Exemple de règles d'arité 2 pour le cas d'une gammapathie monoclonale à signification indéterminée.

La production des règles d'arité n nécessite l'utilisation du moteur d'interrogation SWRL-IQ (cf. 4.1.3.3).

```
Finding_Diagnosis_Lrplus [IgA_Measurement_Less_Than_3g_Per_dL_1, MGUS_1, context_1, 0.05, 0.25, 0.9, 1.25]
Finding_Diagnosis_Lrminus [IgA_Measurement_Less_Than_3g_Per_dL_1, MGUS_1, context_1, 0.05, 0.25, 0.9, 0.9]
```

Figure 57: Exemple de règles d'arité n pour le cas d'une gammapathie monoclonale à signification indéterminée.

### 5.2.7 La représentation des axiomes de fermeture

Ce processus consiste à ajouter à des définitions existantes, des restrictions avec un quantificateur universel, par exemple, pour la restriction avec quantificateur existentiel suivante : « *Disease\_Excludes\_Finding* **some** *Bone\_Lesion* », nous créons une autre restriction avec un quantificateur universel : « *Disease\_Excludes\_Finding* **only** *Bone\_Lesion* ».

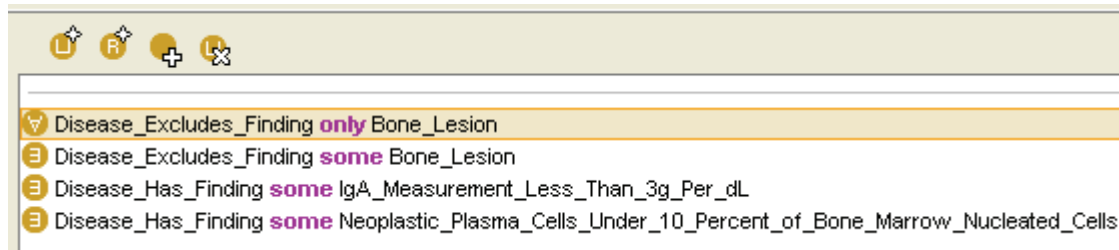


Figure 58: Définition formelle partielle d'une MGUS avec un axiome de fermeture. Cette restriction signifie qu'une MGUS n'exclut que les lésions osseuses.

### 5.2.8 Conclusion

Le modèle de connaissance obtenu après instanciation contient un ensemble de règles d'ordre 0. Ces règles d'ordre 0 permettent ensuite de créer des ensembles de triplets, quintuplets et septuplets (cf. 4.2.7) qui seront stockés dans une base de données relationnelle afin d'être utilisés par un moteur d'inférences.

## 5.3 La description du système

Nous présentons dans cette section le fonctionnement du système. Comme dit plus haut, c'est un système d'aide à la décision à base d'ontologie, c'est-à-dire que la structure et le contenu de toutes les composantes du système sont définis par l'ontologie. Les programmes (code source) **Java** sont présentés en annexe. Le système est accessible via ces liens :

- ❖ <http://nautilus.univ-rennes1.fr:8080/AppDiagnosticRatioWithDetail/Auth.do> pour la version avec moteur de raisonnement abductif-ratio. Les diagnostics sont triés via des ratios. Ce ratio est le quotient du nombre de signes observés et le nombre de signes caractérisant la maladie concernée.
- ❖ <http://nautilus.univ-rennes1.fr:8080/AppDiagnosticVdpWithDetail/Auth.do> pour la version avec moteur de raisonnement abductif-probabiliste. Les diagnostics sont triés via les probabilités des diagnostics sur la base des signes observés.

### 5.3.1 La création des ensembles et stockage dans une base de données relationnelle

Les APIS (JESS, SWRL et SWRL-IQ) permettent de produire des ensembles (Tableau 11, Tableau 12, Tableau 13 et Tableau 14) à partir de la base de règles d'ordre 0. Ensuite, ces ensembles sont stockés dans une base de données relationnelle (Figure 28).

Tableau 11: Vue partielle de l'ensemble des triplets (r,f,d).

| Relation (r)                       | Finding (f)                       | Disease_And_Disorders (d) |
|------------------------------------|-----------------------------------|---------------------------|
| Finding_Has_Diagnosis              | IgA_Measurement_Under_3g_Per_dl_1 | MGUS_1                    |
| Finding_Absence_Excludes_Diagnosis | IgA_Measurement_Under_3g_Per_dL_1 | MGUS_1                    |
| Finding_Excludes_Diagnosis         | Bone_Lesion_1                     | MGUS_1                    |
| Finding_Absence_Has_Diagnosis      | Bone_Lesion_1                     | MGUS_1                    |

Tableau 12: Vue partielle de l'ensemble des quintuplets (r,f,d,tv,u).

| Relation (r)          | Finding (f)       | Disease_And_Disorders (d) | Threshold Value (tv) | Unit (u) |
|-----------------------|-------------------|---------------------------|----------------------|----------|
| Finding_Has_Diagnosis | IgA_Measurement_1 | MGUS_1                    | < 3                  | g/dL     |
| Finding_Has_Diagnosis | IgA_Measurement_1 | MM_1                      | ≥ 3                  | g/dL     |

Tableau 13: Vue partielle de l'ensemble des septuplets (f,d,c,prev,Se,Sp,LrMinus).

| Finding (f)                       | Disease_And_Disorders (d) | Context (c) | Prévalence (prev) | Se   | Sp  | LrMinus |
|-----------------------------------|---------------------------|-------------|-------------------|------|-----|---------|
| IgA_Measurement_Under_3g_Per_dl_1 | MGUS_1                    | context_1   | 0.05              | 0.25 | 0.9 | 0.9     |

Tableau 14: Vue partielle de l'ensemble des septuplets (f,d,c,prev,Se,Sp,LrPlus).

| Finding (f)                       | Disease_And_Disorders (d) | Context (c) | Prévalence (prev) | Se   | Sp  | LrPlus |
|-----------------------------------|---------------------------|-------------|-------------------|------|-----|--------|
| IgA_Measurement_Under_3g_Per_dl_1 | MGUS_1                    | context_1   | 0.05              | 0.25 | 0.9 | 1.25   |

Ces ensembles sont ensuite utilisés par le moteur d'inférences pour les tâches de raisonnement. Pour les tests et les évaluations du système, nous n'avons utilisé que les connaissances à caractères dichotomiques parce que les dossiers qu'on avait à notre disposition ne contenaient que des données de cette nature.

### 5.3.2 Le moteur d'inférences mixte

Il s'agit d'un moteur (cf. 4.2.8) implémentant les modes de raisonnement abductif (Walton, 2004) (cf. 3.3.2.7) et probabiliste. Les signes observés par le praticien (a) sont transmis au module de raisonnement « abductif », ce dernier interroge la base de connaissances et extrait toutes les hypothèses de maladies (diagnostic) qui correspondent aux signes observés. Ensuite, il crée une liste des diagnostics possibles et une liste des diagnostics exclus. Chaque diagnostic est chargé dans la liste correspondante avec ses signes (présents ou absents). Si les probabilités sont disponibles alors il dérive les probabilités des diagnostics et trie la liste des diagnostics possibles uniquement via ces probabilités. Sinon, si les probabilités ne sont pas disponibles alors il calcule les ratios et trie la liste des diagnostics possibles via ces ratios. Le logigramme<sup>23</sup> (Figure 59) suivant présente les étapes de traitement des données dans le moteur d'inférences.

<sup>23</sup> Outil permettant de décrire et de visualiser de façon séquentielle et logique les actions à mener et les décisions à prendre pour atteindre un objectif. [Wikipédia]

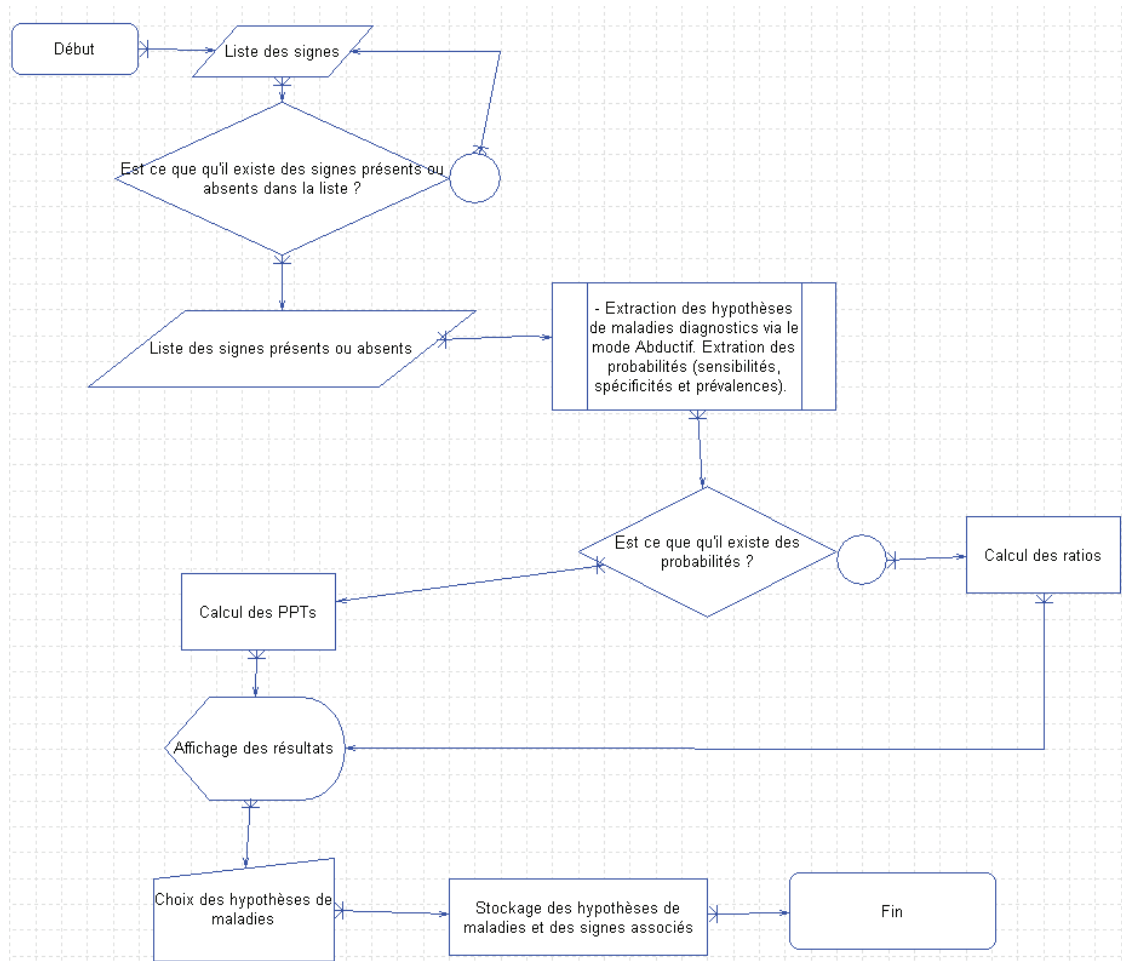


Figure 59: Logigramme du moteur d'inférences mettant en œuvre les modes de raisonnement abductif et probabiliste.

### 5.3.3 L'interface du système

ODES propose de manière standard quatre formulaires : un formulaire de connexion, un formulaire de choix de l'ontologie (base de connaissances) à utiliser, un formulaire adaptatif contenant tous les signes et symptômes groupés par famille (signes cliniques, signes radiologiques, signes biologiques, etc.) et un formulaire de résultat présentant les diagnostics possibles et les diagnostics exclus.

#### 5.3.3.1 Le formulaire de connexion

ODES est conçu pour être exploité par plusieurs utilisateurs et ce formulaire permet de différencier ces utilisateurs.

Le formulaire de connexion, intitulé "User Login", se trouve dans une fenêtre de navigateur. Il contient deux champs de saisie : "Utilisateur" avec la valeur "dregis" et "Mot de passe" avec la valeur "9999". Un bouton "Login" est situé en dessous. Une note en bas du formulaire indique : "Log in = dregis and password = U936 and press OK for enter".

Figure 60: Formulaire de connexion au système. Une fois l'utilisateur authentifié toutes les interrogations et décisions de ce dernier seront stockées avec son identifiant.

### 5.3.3.2 Le formulaire de choix de l'ontologie

Ce formulaire de choix permet de choisir l'ontologie ou la base de connaissances que nous souhaitons utiliser.

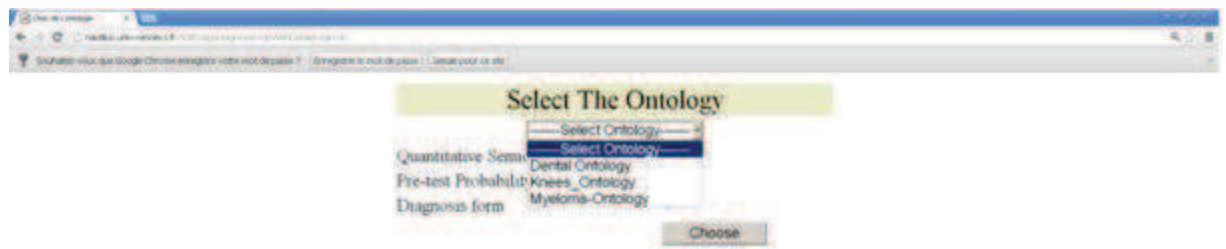


Figure 61: Formulaire de choix de l'ontologie ou de la base de connaissances à utiliser.

### 5.3.3.3 Le formulaire adaptatif

Le formulaire adaptatif ou d'interrogation (Frank & Szekely, 1998) (Figure 63) du système est fonction du contenu de l'ontologie. Ces signes sont extraits à partir des requêtes listées ci-après (Figure 62). Les signes sont regroupés par catégorie, par exemple biologiques, radiologiques, symptômes, etc. Ces requêtes (Figure 62) permettent de créer des ensembles (signe, catégorie) qui sont stockés dans deux tables « E\_Finding et E\_Category » de la base de données (Figure 28).

| Name                      |   |
|---------------------------|---|
| List_Lesion               | → Lesion(?)! → sqwrl:select(?)!             |
| List_Radiological_Finding | → Radiologic_Finding(?)r → sqwrl:select(?)r |
| List_Symptom              | → Symptom(?)s → sqwrl:select(?)s            |
| List_Test_Result          | → Test_Result(?)t → sqwrl:select(?)t        |

Figure 62 : Les requêtes de sélection des catégories de signes. Ces catégories sont les superclasses des classes (signes) de l'ontologie. Elles peuvent contenir d'autres catégories, par exemple la catégorie « Lesion » contient les catégories « Bone\_Lesion », « Lytic\_lesion », etc.

Le contenu de la table « E\_Category » est ensuite utilisé pour définir les cadres (groupes) dans lesquels les signes « E\_Finding » seront groupés sur le formulaire (Figure 63).

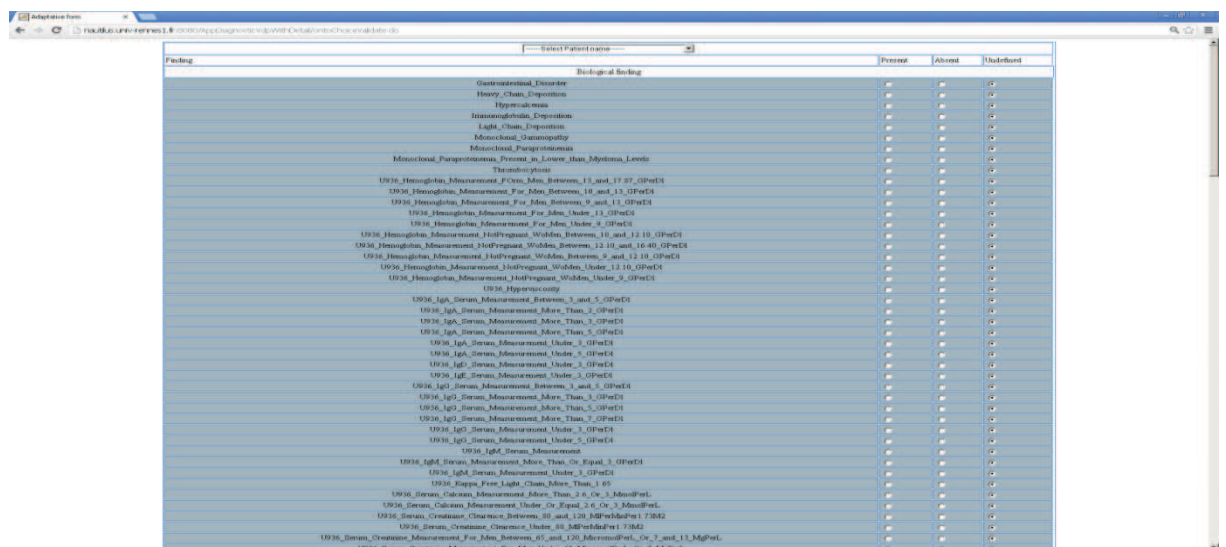


Figure 63: Formulaire adaptatif ou d'interrogation, ce formulaire est fonction de chaque base de connaissances ou ontologie.



### 5.3.3.4 Le formulaire de résultat

Le formulaire des résultats (Figure 64) affiche la liste des diagnostics possibles et celle des diagnostics exclus. Pour chaque diagnostic, la liste des signes présents et absents est fournie d'une part, et d'autre part, la liste des signes non observés est aussi fournie. Le nombre de signe observé est aussi affiché, ainsi que le ratio. Pour les cas avec des probabilités, à la place du ratio c'est la probabilité post-test qui est affichée. Les cases à cocher (situées après le ratio) permettent de sélectionner les diagnostics retenus par le praticien. Ces derniers sont stockés dans la base de cas avec les signes présents et absents qui y sont rattachés. Les diagnostics proposés par le système comme « possible » ou même comme « exclus » sont aussi stockés dans la base des cas avec les signes qui y sont rattachés. Ensuite à partir de ces données stockées dans la base cas, on peut calculer les sensibilités et spécificités des signes pour les maladies et les prévalences des maladies (cf. 5.4).

The screenshot shows a web application interface for 'Diagnostic Result'. At the top, there's a 'Select Disease name' dropdown. Below it, four diagnostic categories are listed, each with a table of findings and their roles. The categories are:

- Monoclonal Immunoglobulin Deposition Disease**: Observed finding: 3, All Finding: 4, Ratio: 0.75. Findings include Neoplastic Plasma Cells Present in Bone Marrow, Immunoglobulin Deposition, Light Chain Deposition, and Heavy Chain Deposition.
- Heavy Chain Deposition Disease**: Observed finding: 3, All Finding: 8, Ratio: 0.375. Findings include Neoplastic Plasma Cells Present in Bone Marrow, Immunoglobulin Deposition, Light Chain Deposition, Heavy Chain Deposition, Amorphous Eosinophilic and Acellular Deposit, Microaneurysm, Monoclonal Gammopathy, and Unfavorable Clinical Outcome.
- Light Chain Deposition Disease**: Observed finding: 3, All Finding: 8, Ratio: 0.375. Findings include Neoplastic Plasma Cells Present in Bone Marrow, Immunoglobulin Deposition, Light Chain Deposition, Heavy Chain Deposition, Amorphous Eosinophilic and Acellular Deposit, Microaneurysm, Monoclonal Gammopathy, and Unfavorable Clinical Outcome.
- Primary Amyloidosis**: Observed finding: 4, All Finding: 30, Ratio: 0.1333. Findings include Gastrointestinal Disorder, U936 Parathyroid, Endocrine Pancreas Disorder, Pancreatic Disorder, U936 Gonadal, Malabsorption Syndrome, and Intestinal Disorder.

Figure 64: Formulaire des résultats.

## 5.4 Le calcul de la sémiologie quantitative

Le jeu de données utilisé pour ces calculs est celui des urgences odontologiques. Les formules utilisées sont présentes dans cette section 2.4. La Figure 65 présente les prévalences produites, la Figure 66 montre les effectifs utilisés pour le calcul des sensibilités et des spécificités des signes pour les maladies et la Figure 67 présente un extrait de ces sensibilités et spécificités des signes pour les maladies.

| Pathologies                         | Prévalence |
|-------------------------------------|------------|
| Pulpite irréversible                | 22,66      |
| Nécrose pulpaire                    | 12,31      |
| Abcès apical aigu                   | 11,33      |
| Abcès Paro aigu                     | 6,89       |
| Syndrome du septum                  | 6,40       |
| Péricoronarite aiguë                | 5,41       |
| Cellulite séreuse                   | 3,94       |
| Abcès péri-coronaire                | 2,95       |
| Cellulite suppurative               | 2,95       |
| Lésion traumatique muqueuse         | 2,46       |
| Fracture coronaire pénétrante       | 1,97       |
| Pulpite subaiguë                    | 1,97       |
| Gingivite aiguë                     | 1,97       |
| Fracture dentaire                   | 1,97       |
| Hyperhémie pulpaire                 | 1,97       |
| Fracture radiculaire pénétrante     | 1,47       |
| Alvéolite suppurative               | 1,47       |
| Cellulite post-chirurgicale         | 0,98       |
| Hémorragie postopératoire           | 0,98       |
| Désordre d'éruption                 | 0,49       |
| Lésion allergique muqueuse          | 0,49       |
| Expulsion dentaire                  | 0,49       |
| Parotidite aiguë                    | 0,49       |
| Lésion endo-paro (O endo)           | 0,49       |
| Abcès par fermeture poche           | 0,49       |
| GUNA                                | 0,49       |
| Intrusion dentaire partielle        | 0,49       |
| Fracture radiculaire non pénétrante | 0,49       |

Figure 65: Prévalences calculées à partir de la base de cas des urgences odontologiques.

| FN | FP | VN | VP | Disease_External_Name | Finding_External_Name           | Prevalence |
|----|----|----|----|-----------------------|---------------------------------|------------|
| 0  | 16 | 0  | 28 | Acute_apical_abcess   | Insomnia                        | 11.3300    |
| 0  | 28 | 0  | 36 | Acute_apical_abcess   | Localized_pain                  | 11.3300    |
| 0  | 5  | 0  | 0  | Acute_apical_abcess   | Mucous_membrane_swelling        | 11.3300    |
| 0  | 5  | 0  | 0  | Acute_apical_abcess   | Neck_haemorrhage                | 11.3300    |
| 0  | 2  | 0  | 4  | Acute_apical_abcess   | Pain_in_decubitus               | 11.3300    |
| 0  | 26 | 0  | 28 | Acute_apical_abcess   | Pain_on_apical_palpation        | 11.3300    |
| 0  | 19 | 0  | 32 | Acute_apical_abcess   | Pain_on_axial_percussion        | 11.3300    |
| 0  | 26 | 0  | 28 | Acute_apical_abcess   | Pain_on_cervical_palpation      | 11.3300    |
| 0  | 25 | 0  | 0  | Acute_apical_abcess   | Pain_on_palpation               | 11.3300    |
| 0  | 44 | 0  | 0  | Acute_apical_abcess   | Pain_on_tooth_percussion        | 11.3300    |
| 0  | 13 | 0  | 15 | Acute_apical_abcess   | Pain_on_transversal__percussion | 11.3300    |
| 0  | 66 | 0  | 30 | Acute_apical_abcess   | Pain_provoked_by_chewing        | 11.3300    |
| 0  | 34 | 0  | 0  | Acute_apical_abcess   | Pain_provoked_by_hot            | 11.3300    |
| 0  | 4  | 0  | 0  | Acute_apical_abcess   | Pain_radiating_to_ear           | 11.3300    |
| 0  | 16 | 0  | 0  | Acute_apical_abcess   | Periodontal_pocket              | 11.3300    |

Figure 66: Extrait des effectifs pour le calcul des sensibilités et des spécificités des signes pour les maladies.

| Disease_External_Name    | Finding_External_Name                                  | Prevalence | Sensitivity | Specificity |
|--------------------------|--|------------|-------------|-------------|
| Acute_periodontal_abcess | Exacerbated_response_to_cold_dental_pulp_vitality_test | 6.8966     | 0.0000      | 5.1282      |
| Irreversible_pulpitis    | Exacerbated_response_to_cold_dental_pulp_vitality_test | 22.6601    | 100.0000    | 40.0000     |
| Acute_apical_abcess      | Abnormal_tooth_mobility                                | 11.3300    | 0.0000      | 4.0000      |
| Irreversible_pulpitis    | Abnormal_tooth_mobility                                | 22.6601    | 0.0000      | 4.5455      |
| Irreversible_pulpitis    | Asthenia   | 22.6601    | 100.0000    | 0.0000      |
| Acute_periodontal_abcess | Periodontal_pocket                                     | 6.8966     | 100.0000    | 0.0000      |
| Acute_periodontal_abcess | Spontaneous_pain                                       | 6.8966     | 100.0000    | 0.0000      |
| Necrosis_of_the_pulp     | Pain_provoked_by_hot                                   | 12.3153    | 100.0000    | 0.0000      |
| Acute_apical_abcess      | Exacerbated_response_to_cold_dental_pulp_vitality_test | 11.3300    | 0.0000      | 100.0000    |

Figure 67: Extrait des sensibilités et des spécificités calculées pour différentes maladies.



## 5.5 Évaluation

L'évaluation du système s'est déroulée à trois endroits différents : (a) l'unité de recherche médicale UMR936, (b) le service des urgences odontologiques du CHU de Rennes et (c) le service de la médecine du sport du CHU de Rennes. Tous les dossiers patients sont anonymisés.

Pour le système, un diagnostic correct est, si les probabilités sont absentes, un diagnostic présent dans la liste des diagnostics possibles renvoyés quelle que soit sa position dans cette liste. Et, si nous avons des probabilités, un diagnostic correct est le premier dans la liste des diagnostics possibles. Les équations et les abréviations utilisées sont décrites dans cette section 4.2.10.2.

### 5.5.1 Matériels

Le premier jeu de données concerne 63 dossiers patients obtenus auprès du service de médecine interne de l'hôpital Sud de Rennes, pour le deuxième jeu de donnée, nous avons 202 dossiers patients provenant du service des urgences odontologiques du CHU de Rennes et 20 dossiers patients fournis par le service de la médecine du sport du CHU de Rennes.

Les ensembles utilisés (cf. 5.3.1) sont produits à partir de règles d'ordre 0 produites sur la base de trois ontologies :

- ❖ L'ontologie concernant les maladies plasmocytaires contient 277 classes, 822 restrictions, 30 pathologies et elle a permis de produire 1710 règles d'ordre 0 (Donfack Guefack, Bertaud-Gounot, Duvauferrier, et al., 2012).
- ❖ Celle des pathologies en urgence odontologique contient 446 classes, 705 restrictions, 113 pathologies, elle a permis de produire 2792 règles d'ordre 0 (Donfack Guefack, Serrand-Obry, Lasbleiz, Duvauferrier, & Bertaud-Gounot, 2012). Et,
- ❖ Celle concernant les lésions traumatiques du genou a permis de produire 130 règles d'ordre 0 (Donfack Guefack, Bertaud-Gounot, Burgun, et al., 2012).

### 5.5.2 Méthodes

#### 5.5.2.1 Jeu de données 1

L'objectif est de comparer les résultats obtenus par le système aux résultats inscrits dans les dossiers. Les résultats contenus dans les dossiers des patients représentent « la référence ». Ces résultats ont été obtenus après avoir effectué des examens complémentaires pour confirmer ou infirmer les hypothèses initialement produites par le clinicien sénior.

#### 5.5.2.2 Jeu de données 2

L'évaluation a été faite en s'appuyant sur les signes observés en direct sur le patient se présentant au service des urgences odontologiques. Les résultats du clinicien sénior étaient confrontés à ceux du système.

#### 5.5.2.3 Jeu de données 3

L'évaluation a été faite en s'appuyant sur les résultats consignés dans les dossiers patients. Ces résultats proviennent de l'IRM (Imagerie par résonance Magnétique) et ils sont considérés comme le Gold Standards. Ensuite nous avons sur la base des signes observés par

un clinicien sénior récupéré les résultats de ce clinicien d'une part et d'autre part nous avons récupéré les résultats du système.

### **5.5.3 Résultats**

#### **5.5.3.1 Jeu de données 1 : 63 dossiers patients du service de médecine interne de l'hôpital Sud de Rennes**

Avant les examens complémentaires les résultats du clinicien sénior étaient : 40 cas de résultats corrects, et 23 cas de résultats incorrects, soit un taux de diagnostics corrects de 63,5% et un taux de diagnostics incorrects de 36,5%. Avec le système et sur la base des mêmes signes observés initialement par le clinicien sénior nous avons obtenu 57 cas de résultats corrects, et 9 cas de résultats incorrects, soit un RCD de 90% et un ER de 10% (Donfack Guefack, Bertaud-Gounot, Duvauferrier, et al., 2012).

#### **5.5.3.2 Jeu de données 2 : 202 dossiers au service des urgences odontologiques du CHU de Rennes**

Les résultats du clinicien étaient comme suit : 150 cas de bons résultats, et 52 cas de mauvais résultats, soit un taux de diagnostics corrects de 74,25 % et un taux de diagnostics incorrects de 25,75 %. Avec le système et sur la base des mêmes signes observés par le clinicien sénior nous avons obtenu 198 cas de résultats corrects, et 4 cas de résultats incorrects, soit un RCD de 98,02% et un ER de 1,98% (Donfack Guefack, Serrand-Obry, et al., 2012).

#### **5.5.3.3 Jeu de données 3 : 20 dossiers patients du service de la médecine du sport du CHU de Rennes**

Les résultats du clinicien se présentent comme suit : 13 cas de bons résultats et 7 cas de mauvais résultats, soit un taux de diagnostics corrects de 65% et un taux de diagnostics incorrects de 35%. Les résultats du système se présentent comme suit : 16 cas de diagnostics corrects, 3 cas de diagnostics non discriminant et 1 cas de diagnostic incorrect, soit un RCD de 80% et un taux de ER de 5% et un taux de diagnostics non discriminants (RDnD) de 15% (Donfack Guefack, Bertaud-Gounot, Burgun, et al., 2012).

## **5.6 Conclusion**

Nous avons donc réalisé et évalué des méthodes de modélisation des signes dans les ontologies à des fins de diagnostic médical. Elles prévoient des moyens de modélisation des connaissances de raisonnement ainsi que la définition formelle des connaissances sur les maladies. Ces méthodes permettent d'exprimer dans un modèle ontologique les connaissances : (a) dichotomiques, (b) multivaluées, (c) probabilistes, (d) sur les définitions multi-niveaux des maladies, (e) sur les signes absents et, (f) sur le raisonnement (de contrôle).

## Chapitre 6

### Discussion et Conclusion

#### 6.1 Introduction

Les systèmes informatiques permettent l'amélioration de la qualité des soins et de la prise en charge des malades par la réduction des erreurs médicales et par la détection des omissions (Seroussi et al., 2013; Zhang et al., 2005). Notre objectif était la mise en place de méthodes générales pour la modélisation des signes dans les ontologies biomédicales pour l'aide au diagnostic et la proposition d'un moteur d'inférences mettant en œuvre un mode de raisonnement adapté au diagnostic médical. Une ontologie est constituée d'un large réseau de concepts d'un domaine liés par des relations et qui peut être exploitée par des systèmes d'aide au diagnostic médical. Cependant, il existe des problèmes :

- ❖ liés au contenu des ontologies biomédicales.
- ❖ liés à la capacité des langages de représentation des connaissances associés à mettre en œuvre le diagnostic médical.

Nous avons proposé des méthodes de représentation des connaissances sur les maladies et des méthodes de représentation des connaissances de raisonnement. Ces méthodes ont permis de produire des modèles de connaissances contenant tous les types (déclaratif, procédural et contrôle) de connaissances. Ces méthodes permettent aussi de représenter ces types de connaissances quel que soit le caractère de ces dernières (dichotomique, multivaluée, probabiliste, multi-niveau). Ces méthodes nous ont permis aussi de mettre en œuvre les raisonnements abductif et probabiliste.

Le système ODES élaboré s'appuie sur le contenu des modèles ontologiques mis en place via ces méthodes et les résultats des tests et évaluations montrent que l'utilisation d'un tel système apporte une aide substantielle aux médecins.

#### 6.2 Modèle de connaissances

Les méthodes mises en place permettent d'avoir un modèle unique contenant les connaissances du domaine (connaissances déclaratives et connaissances procédurales) et les connaissances de contrôle (connaissances de raisonnement). L'avantage d'avoir un modèle unique réside sur le fait que les mises à jour du modèle en sont facilitées. Et lors de l'exploitation du modèle, le moteur d'inférences ne traite qu'un seul modèle et non deux comme c'est le cas avec le système ODDIN (García-crespo et al., 2010).

Les méthodes (cf. 4.2.1 et 5.2) élaborées ont permis de compléter (ajouter des connaissances) des ontologies de domaine. Ces méthodes sont génériques car elles ont permis de traiter trois ontologies provenant de trois disciplines différentes. Les modèles ontologiques résultants contiennent des connaissances du domaine et les connaissances de contrôle qui sont

clairement séparées. Les connaissances du domaine sont formalisées par (a) une représentation abstraite (sans tenir compte de l'utilisation) et hiérarchique des concepts du domaine, et (b) une représentation des relations maladie-sign(e)s. Ces modèles sont théoriques, contrairement au modèle à base de règles qui n'expriment que les connaissances pratiques de l'expert (Bruaux, 2007; Page, 1990) (cf. 1.1).

Certains fragments de connaissances procédurales (Figure 48)(cf. 4.2.4, Règle. 11 et 12) sont exprimés via des builtins SWRL (Elenius, 2012; Peter et al., 2004) sans recours aux programmes informatiques complexes comme ce fut le cas pour les modèles à base de *frames* (cf. 1.1) (Grigorova & Nikolov, 2008; Turban & Liang, 2005; Winograd, 1975).

Nous avons intégré des connaissances de contrôle via l'ajout de nouvelles relations (ObjectProperty et DatatypeProperty) et des règles SWRL dans nos modèles ontologiques. Cette intégration de nouvelles connaissances via ces règles et relations permet de séparer les connaissances sur les maladies (connaissances du domaine) et les connaissances de contrôle (connaissances de raisonnement), contrairement au modèle à base de règles qui compile toutes ces connaissances dans leurs règles de production (Bruaux, 2007; Page, 1990) (cf. 1.1). Les connaissances de contrôle sont exprimées de manière abstraite en utilisant les classes et les relations de l'ontologie (Clancey, 1983). Cette abstraction donne la capacité aux modèles d'être réutilisables. Aussi, ces connaissances de contrôle permettent au système de fournir des explications plus précises sur les inférences (Bruaux, 2007; Page, 1990). Le système peut à ce moment fournir des informations sur le « POURQUOI » il a évoqué une hypothèse de maladie, « COMMENT » il a confirmé ou infirmé cette hypothèse, contrairement au système exploitant des modèles à base de règles qui ne peuvent qu'expliquer POURQUOI, COMMENT, voire POURQUOI PAS une règle a été utilisée (Bruaux, 2007; Clancey, 1981) (cf. 1.1).

Nos modèles ontologiques sont des modèles hiérarchiques, c'est-à-dire qu'ils sont organisés sous forme de taxonomie avec des relations de généralisation et de spécialisation (par exemple, un myélome multiple est une maladie plasmocytaire). A ces relations s'ajoutent des relations d'association (par exemple, une maladie « d » a pour signe « f »). Cette caractéristique (hiérarchique) leur donne la capacité de faciliter l'ajout, la modification et la suppression de concepts ou de règles tout en garantissant la conservation de la cohérence des modèles.

## 6.2.1 Pourquoi une instanciation complète du modèle ontologique ?

### 6.2.1.1 Utilisation des règles SWRL

L'utilisation des règles SWRL et des formalismes d'interrogation SQWRL, SWRL-IQ impose la présence d'instances dans le modèle ontologique car ces formalismes ne fonctionnent qu'en présence d'instances (I. Horrocks et al., 2004). Cette utilisation nous permet de garantir la séparation des définitions des maladies et des définitions utiles pour le processus de raisonnement. En effet, un des rôles majeurs d'une ontologie biomédicale est de servir de vocabulaire, c'est-à-dire fournir une liste des noms d'entité et des relations entre ces entités



(Bodenreider, 2008). Le fait de séparer les définitions des maladies et les définitions utiles pour le processus de raisonnement permet de garantir la conservation de la conformité de ce vocabulaire sur les maladies. Cette approche diffère de celle de Rodríguez-González (Rodríguez-González, Labra-Gayo, et al., 2012) qui propose de modifier les définitions des maladies pour être conforme aux requêtes soumises au système (cf. 3.3.4, Figure 7 et Figure 8). Notre approche permet de garantir cette séparation et permet de tirer parti de ces deux modes de vision des ontologies biomédicales, à savoir comme vocabulaire sur les maladies, et comme source de données (nos ensembles créés à partir de TBox, ABox et les règles) utilisable par les systèmes informatiques pour les tâches d'aide à la décision (Musen, 1992).

### **6.2.1.2 La prise en charge de la conservation de la hiérarchie**

Un des intérêts des modèles ontologiques est la hiérarchie des concepts qu'ils comportent. Cette hiérarchisation des concepts facilite la mise à jour des modèles, c'est-à-dire l'ajout, la modification et la suppression des concepts ou des règles tout en maintenant la cohérence du modèle.

En revanche, lors de l'instanciation, cette prise en compte de la hiérarchie n'est pas tacite, c'est-à-dire qu'elle ne se fait pas de manière automatique, elle doit être explicitement faite. Par exemple, une MGUS exclut la présence des lésions osseuses (Figure 49), les lésions lytiques sont des lésions osseuses (Figure 52), et grâce à cette prise en compte de la hiérarchisation des concepts, une assertion MGUS exclut les lésions lytiques est créée (cf. 5.2.6, Figure 51 et Figure 53)(Donfack Guefack, Bertaud-Gounot, Duvauferrier, et al., 2012).

### **6.2.2 Le raisonnement abductif**

Selon Walton (Walton, 2004) le diagnostic médical est une forme de raisonnement abductif. Ce type de raisonnement donne la possibilité de générer des hypothèses de maladies même si on n'a pas observé tous les signes ou symptômes caractérisant ces hypothèses (Pottier & Planchon, 2011). Les moteurs d'inférences classiques (par exemple Pellet, Racer, etc.) ne gèrent pas ce type de raisonnement car ils mettent en œuvre un raisonnement déductif (Bertaud-Gounot, Duvauferrier, et al., 2012; Le Moing, 2009; Prcela et al., 2008).

Notre proposition s'appuie sur les travaux de Cory Henson (Henson et al., 2011) et elle consiste à produire des ensembles de triplets, quintuplets et septuplets (cf. 5.3.1). Ces ensembles contiennent les relations entre les signes et les maladies (cf. 3.3.4.6) (Figure 68). Ces relations peuvent être déterministes, c'est le cas matérialisé par les triplets et les quintuplets. Elles peuvent être probabilistes, c'est le cas matérialisé par les septuplets. Ces ensembles sont ensuite utilisés par le moteur d'inférences mis en place pour produire les hypothèses de maladies.

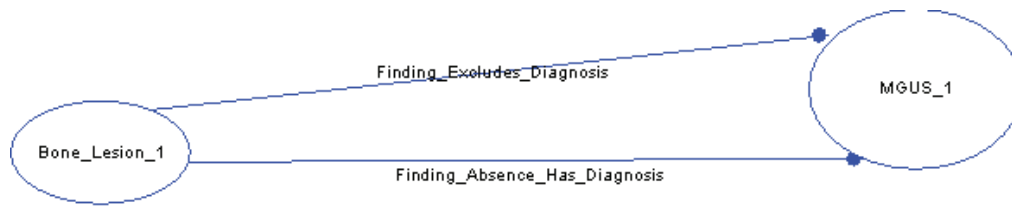


Figure 68: Illustration d'un triplet sous forme d'un réseau de relations signe-diagnostic (hypothèse d'une maladie).

## 6.2.3 Les signes absents

### 6.2.3.1 La représentation

Les signes absents ou négatifs sont gérés de façon efficace par notre approche (cf. 4.2.2). L'approche de Bertaud-Gounot (Bertaud-Gounot, 2008) entraîne la création de nouvelles classes, cela alourdit considérablement le modèle. Par exemple concernant le signe « Bone\_Lesion » pour une MGUS, on créerait une autre classe « No\_Bone\_Lesion » afin d'obtenir «MGUS » *Has\_Finding* «No\_Bone\_Lesion ». En revanche, notre approche, similaire à celle de Werner Ceusters (Ceusters et al., 2006, 2007) nous permet de n'avoir qu'un seul signe « Bone\_Lesion » et une relation du type « Excludes » (Figure 49). Cette approche permet de traduire le fait qu'une maladie exclut la possibilité ou l'éventualité d'un signe, ou qu'un signe exclut la possibilité ou l'éventualité d'une maladie ou hypothèse de maladie (Figure 68).

Un autre intérêt de la représentation des signes absents est qu'ils permettent de produire des effectifs (par exemple le nombre de faux négatifs et le nombre de vrais négatifs, cf. 5.4 pour des détails) qui sont utilisés pour la production des sensibilités et des spécificités des signes pour les maladies concernées (cf. 6.2.4.2).

### 6.2.3.2 Le raisonnement abductif et les signes absents

L'utilisation des ensembles (cf. 5.3.1) conçu à partir des règles (cf. 4.2.3) permet de mettre en œuvre le raisonnement abductif et de prendre en charge les signes absents ou négatifs (cf. 4.2.2) (Bertaud-Gounot, Donfack Guefack, Lasbleiz, Bourdé, & Duvauferrier, 2011; Donfack Guefack, Bertaud-Gounot, Duvauferrier, et al., 2012). Les règles 1,2,3,4,5,6,7,8,9,10,11 et 12 (cf. 4.2.3) qui ont servi à la production des ensembles ont été conçues en nous appuyant sur les stratégies de raisonnement (Arocha et al., 2005) des praticiens (cliniciens, médecins, etc.). Ces ensembles permettent de construire un réseau de relations entre les signes et les hypothèses de maladie (Figure 68).

## 6.2.4 Ontologie et Probabilité

### 6.2.4.1 Représentation

La représentation des probabilités (sensibilités et spécificités des signes pour les maladies, et les prévalences des maladies pour une population donnée) a été faite via la réification des relations en utilisant l'Ontology Design Pattern 1 (ODP) du W3C (Hayes et al., 2006). Notre approche présente plusieurs avantages :

- ❖ Nous utilisons une ontologie existante, le langage de représentation OWL recommandé par le W3C et le moteur d'inférences Pellet. Les définitions existantes sont conservées, car notre objectif est d'avoir dans le même modèle les définitions des maladies telles qu'elles sont dans la littérature et les connaissances de raisonnement. Elle détonne de l'approche proposée par Fenz (Fenz, 2012) sur le fait que nous faisons cohabiter un modèle déterministe et un modèle probabiliste, alors que celle de Fenz consiste à remplacer le modèle ontologique par un réseau bayésien (RB).
- ❖ Les relations (maladie-signes) sont quantifiées contrairement à l'approche de Bertaud-Gounot (Bertaud-Gounot, 2008) qui opte pour l'ajout de relations « diagnoses et justify » avec des facteurs de certitude. Dans cette optique son approche de représentation est qualitative. Les probabilités dans notre approche sont intégrées dans l'ontologie via le processus de réification (cf. 5.2.3) en utilisant des restrictions, ce qui nous donne la possibilité de vérifier facilement la cohérence du modèle, alors que García-crespo (García-crespo et al., 2010) propose l'utilisation d'une ontologie et un fichier texte externe contenant les probabilités. L'exploitation de deux modèles distincts rend les mises à jour délicates et très difficiles.

#### 6.2.4.2 Le raisonnement probabiliste

Notre ontologie probabiliste obtenue contient les couples (maladie, signe) probabilisés et les couples (maladie, contexte) probabilisés. Les probabilités des couples (maladie, signe) sont les sensibilités et spécificités des signes pour les maladies. Les probabilités des couples (maladie, contexte) sont les prévalences des maladies pour une population donnée (contexte). Ces connaissances sont utilisées lors des processus de raisonnement (cf. 2.4) pour dériver les valeurs prédictives des maladies sur la base des signes observés par le médecin. Les règles de Bayes (Aronoff, 2010; Attia, 2003; McGee, 2002) exploitées dans notre approche sont formellement établies, contrairement aux règles de calcul tirées du système MYCIN (Buchanan & Shortliffe, 1984) utilisées par Bertaud-Gounot (Bertaud-Gounot, 2008) qui ne reposent sur aucune base théorique formelle (Heckerman, Horvitz, & Nathwani, 1992; Lucas, 2001).

#### 6.2.5 Les définitions multi-niveaux

Notre approche permet de gérer les définitions multi-niveaux (cf. 3.3.4, Figure 4 et Figure 6) et de raisonner sur ces définitions. Deux propositions ont été faites, l'une s'appuyant sur la propagation transitive (Seidenberg & Rector, 2006) matérialisée par le chaînage des relations (cf. 3.3.2.3, 5.2.4.1 et Figure 47). L'autre s'appuyant sur l'instanciation (Figure 54 et Figure 55). Ces deux propositions permettent de conserver la profondeur des définitions des maladies présentes dans l'ontologie (cf. 5.2.6.2). Elles présentent l'avantage de ne pas requérir des modifications sur la description des maladies pour pouvoir faire du diagnostic médical dessus comme a fait Rodríguez-González (Rodríguez-González, Labra-Gayo, et al., 2012).

## 6.3 Système ODES

### 6.3.1 Évaluation

L'évaluation du système a permis de ressortir les informations suivantes :

- ❖ pour le premier jeu d'essai les 10% d'erreurs sont dues à l'absence de certaines connaissances dans l'ontologie, car après intégration de ces connaissances manquantes, nous avons eu 100% de résultats corrects.
- ❖ pour le deuxième jeu d'essai, les 1,98% d'erreurs sont aussi dues à des erreurs de représentations des connaissances dans l'ontologie et,
- ❖ pour le troisième jeu d'essai, les 5% d'erreurs et les 15% de diagnostics non discriminants sont dus à la qualité des probabilités.

Ces résultats montrent l'utilité d'un tel système dans l'activité des médecins. Cependant, nous n'avons pas mis un accent particulier sur la convivialité de l'interface utilisateur. En effet, un des objectifs de ce dernier (ODES) était de montrer la faisabilité de notre approche.

### 6.3.2 Temps de traitement

Les temps de réponse lors des cycles de raisonnement sont inférieurs à 1 seconde pour des ontologies ayant plus de 30 maladies et plus de 100 signes et symptômes (cf. 4.1.1.2, 4.1.1.3, 4.1.1.4 et 5.5.1), contrairement aux approches proposées par Rodríguez-González (Rodríguez-González & Alor-Hernández, 2013; Rodríguez-González, Torres-Niño, et al., 2012; Rodríguez-González, Labra-Gayo, et al., 2012) qui ont des temps de réponses importants (cf. 3.3.4) pour des ontologies relativement petites (5 maladies et 11 signes).

### 6.3.3 Calcul de la sémiologie quantitative

Un des avantages du système ODES est sa base de cas, car les décisions des médecins qui y sont stockées peuvent ensuite servir pour le calcul des probabilités. A partir de ces données (les diagnostics possibles, exclus et retenus ainsi que leurs signes présents et absents) on peut calculer les effectifs (vrai positifs, faux positifs, vrai négatifs et faux négatifs). Ensuite, avec ces effectifs, on peut produire les sensibilités (Se) et spécificités (Sp) des signes pour les maladies (diagnostics) et les prévalences (Prev) des maladies pour la population considérée (notre base de cas).

Une analyse des données extraites et qui ont été utilisées pour le calcul des Se, Sp et Prev montre qu'il y a un nombre négligeable de vrai négatifs. A titre de rappel, un vrai négatif est caractérisé par l'absence de la maladie et du signe. Ce faible nombre s'explique par une négligence de l'investigateur lors du remplissage du formulaire. En effet, le praticien s'est parfois exclusivement concentré sur les signes présents, négligeant de signaler les signes absents. Or, certaines pathologies sont définies par l'absence de certains signes (Ceusters et al., 2006, 2007), par exemple, le cas d'une MGUS et les critères CRAB. Le tableau ci-dessous montre une partie des « vrais négatifs » observés :

| FN | FP | VN | VP | Disease_External_Name | Finding_External_Name                                  | Prevalence |
|----|----|----|----|-----------------------|--|------------|
| 0  | 24 | 1  | 0  | Acute_apical_abcess   | Abnormal_tooth_mobility                                | 11.3300    |
| 0  | 0  | 3  | 0  | Acute_apical_abcess   | Exacerbated_response_to_cold_dental_pulp_vitality_test | 11.3300    |
| 0  | 21 | 1  | 0  | Irreversible_pulpitis | Abnormal_tooth_mobility                                | 22.6601    |
| 0  | 3  | 2  | 37 | Irreversible_pulpitis | Exacerbated_response_to_cold_dental_pulp_vitality_test | 22.6601    |

Figure 69: Tableau des effectifs utilisés pour le calcul des sensibilités et des spécificités des signes pour les maladies.

Cette observation montre l'intérêt de la prise en compte des signes absents dans les modèles de connaissances en médecine ainsi que l'intérêt de leur traitement par les systèmes d'aide au diagnostic. Et aussi, cela apporte un éclairage sur le taux d'erreur de 5 % et le taux de diagnostics non discriminants de 15% (cf. 5.5.3.3) rencontrés lors de l'évaluation du système.

## 6.4 Conclusion

Notre avons présenté l'intérêt de notre approche en mettant en emphase la capacité du modèle ontologique à prendre en compte les différents types de connaissances (cf. 3.2) et sa capacité résoudre les problèmes liés au diagnostic médical. Les résultats de l'évaluation montrent que l'exploitation d'un système utilisant ce type de modèle de connaissance permet d'améliorer les performances des praticiens.

Nous avons élaboré des méthodes pour enrichir une ontologie biomédicale à des fins de diagnostics. Ces méthodes s'appuient sur les formalismes et outils recommandés par le W3C, et elles facilitent la séparation entre les connaissances sur les maladies (connaissances du domaine) et les connaissances de raisonnement dites aussi connaissances de contrôle. Ces méthodes permettent :

- ❖ d'intégrer et de prendre en charge les définitions des maladies multi-niveaux (Rodríguez-González, Labra-Gayo, et al., 2012),
- ❖ d'exprimer les probabilités conditionnelles et les probabilités *a priori* via des restrictions (Hayes et al., 2006),
- ❖ d'exprimer les connaissances imprécises et vagues en nous appuyant sur leurs définitions intrinsèques (Peelen et al., 2007),
- ❖ de gérer les signes négatifs ou absents et (Ceusters et al., 2006, 2007),
- ❖ de fournir les ensembles utiles pour mettre en œuvre les raisonnements abductif (Henson et al., 2011) et probabiliste (Aronoff, 2010; Donfack Guefack, Bertaud-Gounot, Burgun, et al., 2012).

Notre approche a consisté à intégrer des règles pour exprimer les connaissances de contrôle permettant de guider le moteur d'inférences du système mis en place (Bertaud-Gounot et al., 2011; Donfack Guefack, Bertaud-Gounot, Duvauferrier, et al., 2012). Ensuite nous avons instancié intégralement l'ontologie et nous avons synthétisé des ensembles (triplets, quintuplets et septuplets) que nous avons enregistrés dans une base de données relationnelle pour les tests et évaluation des modèles mis en place.

Le système que nous avons mis en place est totalement piloté par le modèle ontologique, les signes contenus sur le formulaire adaptatif ou formulaire de recueil des

signes sont contenus dans l'ontologie. Trois jeux d'essais et trois modèles différents ont été utilisés pour les tests et les évaluations de l'approche.

Les résultats ont été satisfaisants dans l'ensemble et nous avons tiré quelques informations importantes qui serviront pour les futurs travaux, comme par exemple, l'intérêt des signes absents pour la production d'une sémiologie quantitative fiable.

## **6.5 Perspectives**

### **6.5.1 A court terme**

L'évaluation du système ODES doit être effectuée de manière multicentrique avec un nombre plus important de dossiers. Et pour ce faire, on définira des caractéristiques d'interfaces qui contribueraient à réduire la complexité d'utilisation. Cette évaluation à grande échelle permettrait potentiellement de produire des probabilités fiables qui seront utilisées par la suite pour compléter les ontologies correspondantes.

Nous avons dans ce projet évalué le système uniquement sur les paramètres à caractères dichotomiques, il serait envisageable de le faire avec des dossiers contenant des paramètres multivalués.

Nous nous sommes focalisés uniquement sur les diagnostics, il serait envisageable d'intégrer dans ces modèles ontologiques l'approche thérapeutique.

### **6.5.2 A moyen et long terme**

Dans ce projet, nous n'avons pas tenu compte des systèmes d'informations hospitaliers (SIH). En effet, les informations utiles pour le diagnostic (comme par exemple l'historique du patient, les allergies, etc.) sont contenues dans le dossier médical de ce dernier. En connectant le système au SIH, certains paramètres pourraient être automatiquement renseignés. Pour cela, il faudrait prévoir une ontologie (garantir l'interopérabilité sémantique) (Bodenreider, 2008) pour la transcription ou l'interprétation des termes provenant du dossier patient au niveau du SIH avant de les fournir au système d'aide au diagnostic.



## Bibliographie

- Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues , Methodological Variations , and System Approaches. *AI Communications*, 7(1), 39–59.
- Abburu, S. (2012). A Survey on Ontology Reasoners and Comparison. *International Journal of Computer Applications*, 57(17), 33–39.
- Aikins, J. (1980). *Prototypes And Production Rules An Approach To Knowledge Representation For Hypothesis Formation*. Stanford University, Stanford, CA 94305. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.80.532>
- Alexandrakis, M. G., Passam, F. H., Dambaki, C., Pappa, C. A., & Stathopoulos, E. N. (2004). The relation between bone marrow angiogenesis and the proliferation index Ki-67 in multiple myeloma. *Journal of clinical pathology*, 57(8), 856–60.
- Alter, S. (1980). Decision support systems: current practice and continuing challenges. In Addison-Wesley (Ed.), *Decision support systems: current practice and continuing challenges* (1st ed., pp. 1–23). Menlo Park, California: Alter Steven.
- Angtuaco, E. J. C., Fassas, A. B. T., Walker, R., Sethi, R., & Barlogie, B. (2004). Radiology Multiple Myeloma : Clinical. *Radiology*, 231(5), 11–23.
- Antoniou, G., & Van Harmelen, F. (2001). *Web Ontology Language : OWL* (pp. 1–26).
- Arocha, J., Wang, D., & Patel, V. (2005). Identifying reasoning strategies in medical decision making: a methodological guide. *Journal of biomedical informatics*, 38(2), 154–71.
- Aronoff, S. C. (2010). Answering a Question Using a Clinical Prediction Rule. In O. USA (Ed.), *Translational Research and Clinical Practice: Basic Tools for Medical Decision Making and Self-Learning* (1st ed., pp. 50–55). Philadelphia, Pennsylvania, USA: Waldo, Nelson E.
- Attia, J. (2003). Moving beyond sensitivity and specificity: using likelihood ratios to. *Australian prescriber*, 26(5), 111–113.
- Baader, F. (2003). Basic Description Logics. In F. Baader (Ed.), *The description logic handbook* (1st ed., pp. 43–95). Cambridge University Press New York, NY, USA: Cambridge University Press New York, NY, USA.
- Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (2003). *The Description Logic Handbook: Theory, Implementation and Applications*. (F. Baader, D. Calvanese, D. McGuinness, D. Nardi, & P. Patel-Schneider, Eds.) (1st ed., pp. 1–574). Cambridge: Cambridge University Press.
- Bachimont, B. (1992). Eléments pour une méthodologie et une épistémologie du contrôle. In Hermès (Ed.), *Le contrôle dans les systèmes à base de connaissances - Contribution à l'épistémologie de l'intelligence artificielle* (1st ed., pp. 29–66). Paris: Hermès.



## Bibliographie

- Bachimont, B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des. In J. Charlet, M. Zacklad, G. Kassel, & D. Bourigault (Eds.), *Ingénierie des connaissances : Evolutions récentes et nouveaux défis* (Eyrolles., pp. 305–323). Paris.
- Baur, A., Bartl, R., Pellengahr, C., Baltin, V., & Reiser, M. (2004). Neovascularization of bone marrow in patients with diffuse multiple myeloma: a correlative study of magnetic resonance imaging and histopathologic findings. *Cancer*, 101(11), 2599–604.
- Bertaud-Gounot, V. (2008). *Mémoire de thèse d'université : Proposition d'un modèle constructiviste de base de connaissances fondé sur une ontologie des signes permettant les raisonnements abductif et déductif*. Université de Rennes 1. Retrieved from <http://www.theses.fr/2007REN1B105>
- Bertaud-Gounot, V., Donfack Guefack, V., Brilletet, E., & Duvauferrier, R. (2012). Les technologies du web sémantique pour un renouveau des systèmes experts en médecine. Principes, problèmes et propositions à partir de l'exemple du myélome dans le NCI-T. In P. M. Staccini, A. Harmel, S. Darmon, & R. Gouider (Eds.), *Quatorzièmes Journées francophones d'informatique médicale* (Vol. 1, pp. 45–58). Tunis: Springer-Verlag France.
- Bertaud-Gounot, V., Donfack Guefack, V., Lasbleiz, J., Bourdé, A., & Duvauferrier, R. (2011). Creating an ontology driven rules base for an expert system for medical diagnosis. *Studies in Health Technology and Informatics*, 169(1), 714–718.
- Bertaud-Gounot, V., Duvauferrier, R., & Burgun, A. (2012). Ontology and medical diagnosis. *Informatics for health & social care*, 37(2), 51–61.
- Bertaud-Gounot, V., Richard, F., Le Moing, C., & Duvauferrier, R. (2010). Ontology of dental emergencies for diagnostic classification. In *Medinfo* (p. 1482).
- Bishop, B., Kiryakov, A., Ognyanoff, D., Peikov, I., Tashev, Z., & Velkov, R. (2011). OWLIM : A family of scalable semantic repositories. *Semantic Web*, 2(1), 33–42.
- Bobillo, F., Delgado, M., & Gómez-Romero, J. (2008). DeLorean: A Reasoner for Fuzzy OWL 1.1. In F. Bobillo, M. Delgado, & J. Gómez-Romero (Eds.), *Proceedings of the 4th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2008)* (pp. 1–12). Karlsruhe (Germany).
- Bobillo, F., Delgado, M., & Gómez-Romero, J. (2012). DeLorean: A reasoner for fuzzy OWL 2. *Expert Systems with Applications*, 39(1), 258–272.
- Bobillo, F., & Straccia, U. (2008). fuzzyDL: An expressive fuzzy description logic reasoner. In *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)* (pp. 923–930). Ieee. doi:10.1109/FUZZY.2008.4630480
- Bodenreider, O. (2008). Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearbook of medical informatics*, 1(1), 67–79.

## Bibliographie

- Bodenreider, O., & Stevens, R. (2006). Bio-ontologies: current trends and future directions. *Briefings in bioinformatics*, 7(3), 256–74.
- Borst Willem, N. (1997). *Construction of engineering ontologies for Knowledge Sharing and Reuse*. The Dutch Graduate School for Information and Knowledge Systems. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.79.917>
- Bouchon-Meunier, B. (2007). Introduction. In Presses Universitaires de France - PUF (Ed.), *La logique floue* (4th ed., pp. 3–59). Paris: Presses Universitaires de France - PUF.
- Bruaux, S. (2007). *Mémoire de thèse d'université : Vers la construction centrée ontologie de modèles de résolution de problèmes: la méthode ONTOKADS*. Université de Picardie Jules Verne. Retrieved from <http://www.france-bs.com/fr/corps-professoral/bruaux>
- Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. (Bruce. G. Buchanan & E. H. Shortliffe, Eds.) (1st ed., p. 716). Menlo Park, California: Addison-Wesley.
- Burgun, A. (2006). Desiderata for domain reference ontologies in biomedicine. *Journal of biomedical informatics*, 39(3), 307–13.
- Burgun, A., & Bodenreider, O. (2001). Mapping the UMLS Semantic Network into general ontologies. In *Proceedings / AMIA ... Annual Symposium. AMIA Symposium* (pp. 81–5).
- Burgun, A., Bodenreider, O., & Jacquelinet, C. (2005). Issues in the Classification of Disease Instances with Ontologies. *Studies in Health Technology and Informatics*, 116(1), 695–700.
- Cardoso, J. (2007). The Semantic Web Vision : Where are We ? *IEEE Intelligents systems*, 22(5), 84–88.
- Carty, F., Shortt, C. P., Shelly, M. J., Eustace, S. J., & O'Connell, M. J. (2010). Whole-Body Imaging Modalities in Oncology. *Thieme Medical Publishers*, 14(1), 68–85.
- Cesar, P., Costa, G., Laskey, K. B., & Laskey, K. J. (2005). PR-OWL : A Bayesian Ontology Language for the Semantic Web. In *Workshop on Uncertainty Reasoning for the Semantic Web, ISWC 2005* (pp. 1–11).
- Ceusters, W., Barry, S., & Goldberg, L. (2005). A terminological and ontological analysis of the NCI Thesaurus. *Methods of information in medicine*, 44(4), 498–507.
- Ceusters, W., Elkin, P., & Smith, B. (2006). Referent tracking: the problem of negative findings. *Studies in health technology and informatics*, 124(1), 741–6.
- Ceusters, W., Elkin, P., & Smith, B. (2007). Negative findings in electronic health records and biomedical ontologies: a realist approach. *International journal of medical informatics*, 76(3), 326–33.

## Bibliographie

- Charlet, J. (2003, January). *Mémoire HDR : L'ingénierie des connaissances. Développements, Résultats et Perspectives Pour La Gestion Des Connaissances Médicales*. Université Pierre et Marie Curie. Retrieved from <http://tel.archives-ouvertes.fr/tel-00006920>
- Chatellier, G., Ménard, J., & Degoulet, P. (1996). Introduction to quantitative semiology. *La revue du praticien*, 46(3), 291–297.
- Clancey, W. J. (1981). The epistemology of a rule-based expert system —a framework for explanation. *Artificial Intelligence*, 20(3), 215–251.
- Clancey, W. J. (1983). *The Advantages of Abstract Control Knowledge in Expert System Design* (pp. 1–18). Palo Alto. Retrieved from [www.aaai.org/Papers/AAAI/1983/AAAI83-068.pdf](http://www.aaai.org/Papers/AAAI/1983/AAAI83-068.pdf)
- Clancey, W. J. (1993). *The Epistemology of A Rule-Based Expert System : A Framework for Explanation* (pp. 191–204). Stanford, CA, USA. Retrieved from <ftp://reports.stanford.edu/pub/.../CS-TR-81-896.pdf>
- Cléret, M., Le Beux, P., & Le Duff, F. (2001). Les systèmes d'aide à la décision médicale. *Les Cahiers du numérique*, 2(1), 125–154.
- Cuenca, B., Horrocks, I., Motik, B., Parsia, B., Patel-schneider, P., & Sattler, U. (2008). OWL 2 : The Next Step for OWL. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4), 309–322.
- Cuggia, M., Besana, P., & Glasspool, D. (2011). Comparing semi-automatic systems for recruitment of patients to clinical trials. *International journal of medical informatics*, 80(6), 371–88. doi:10.1016/j.ijmedinf.2011.02.003
- Dameron, O., Rubin, D., & Musen, M. (2005). Challenges in Converting Frame-Based Ontology into OWL : the Foundational Model of Anatomy Case-Study Converting the FMA into OWL-DL. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 1(1), 181–185.
- Davis, R., & Buchanan, B. G. (1984). Meta Level Knowledge. In Bruce. G. Buchanan & E. H. Shortliffe (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (1st ed., pp. 507–568). Menlo Park, California: Addison - Wesley.
- De Besoist, B., McLean, E., Egli, I., & Cogswell, M. (2008). *Worldwide prevalence of anaemia 1993-2005 Who Global Database on Anaemia* (pp. 1–6). Atlanta.
- De Dombal, F., Leaper, D., Horrocks, J., Staniland, J., & McCann, P. (1974). Human and Computer-aided Diagnosis of Abdominal Pain: Further Report with Emphasis on Performance of Clinicians. *British medical journal*, 1(5904), 376–380.
- De Dombal, F., Leaper, D., Staniland, J., McCann, P., & Horrocks, J. (1972). Computer-aided diagnosis of acute abdominal pain. *British medical journal*, 2(5804), 9–13.

## Bibliographie

- De Jong, T. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105–113.
- Decaux, O., Cuggia, M., Ruelland, A., Cazalets, C., Cador, B., Jegou, P., & Grosbois, B. (2006). Monoclonal gammopathies of undetermined significance and their progression over time. Retrospective study of 190 patients. *La presse médicale*, 35(7-8), 1143–1150.
- Degoulet, P., & Fieschi, M. (1994a). Raisonnement médical et décision. In Masson (Ed.), *Informatique médicale* (1st ed., pp. 47–63). Paris: MASSON.
- Degoulet, P., & Fieschi, M. (1994b). Données médicales et interprétations. In MASSON (Ed.), *Informatique médicale* (Première., pp. 33–46). Paris: MASSON.
- Désilets, M. (1997). Connaissances déclaratives et procédurales : des confusions à dissiper. *Revue des sciences de l'éducation*, 23(2), 289–308.
- Ding, Z., Peng, Y., & Pan, R. (2006). BayesOWL : Uncertainty Modelling in Semantic Web Ontologies. In Z. Ma (Ed.), *Soft Computing in Ontologies and Semantic Web* (pp. 3–29). Springer-Verlag.
- Dolia, P. (2010). Integrating Ontologies into Multi-Agent Systems Engineering (MaSE) for University Teaching Environment. *Journal of Emerging Technologies in Web Intelligence*, 2(1), 42–47.
- Donfack Guefack, V., Bertaud-Gounot, V., Burgun, A., Barton, A., Duvauferrier, R., & Lasbleiz, J. (2012). Ontologie sémiologique biomédicale et sémiologie quantitative. In *IC pour l'Interopérabilité Sémantique dans les applications en e-Santé, Atelier associé à la conférence IC 2012* (pp. 1–6). Paris: IC'2012.
- Donfack Guefack, V., Bertaud-Gounot, V., Duvauferrier, R., Bourdé, A., Morelli, J., & Lasbleiz, J. (2012). Ontology Driven Decision Support Systems for Medical Diagnosis-An interactive form for consultation in patients with plasma cell disease. *Studies in Health Technology and Informatics*, 180(1), 108–112.
- Donfack Guefack, V., Serrand-Obry, F., Lasbleiz, J., Duvauferrier, R., & Bertaud-Gounot, V. (2012). Système d'aide à la décision à base d'ontologie en urgence odontologique. In *Biologie Santé et Innovation Technologique (BIOSIT)* (Ed.), *Journee Doctorants Post-Doctorants* (p. 1). Rennes: Biologie Santé et Innovation Technologique (BIOSIT).
- Durie, B. (2006). The role of anatomic and functional staging in myeloma: description of Durie/Salmon plus staging system. *European journal of cancer (Oxford, England : 1990)*, 42(11), 1539–43.
- Duron, F. (1999). Hypercalcémie. *Faculté de médecine Pierre et Marie Curie*. Retrieved from [s.jussieu.fr/polys/endocrino/poly/POLY.Chp.7.html](http://s.jussieu.fr/polys/endocrino/poly/POLY.Chp.7.html)
- Eclipse-Foundation. (2012). Eclipse Juno. *The Eclipse Foundation*. Retrieved January 14, 2012, from <http://www.eclipse.org/>

## Bibliographie

- Elenius, D. (2012). SWRL-IQ : A Prolog-based Query Tool for OWL and SWRL. In M. Horridge & P. Klinov (Eds.), *Proceedings of OWL: Experiences and Directions Workshop 2012* (pp. 1–12). Heraklion, Crete, Greece: CEUR-WS.org.
- Elenius, D., & Riehemann, S. (2012). *SWRL-IQ User Manual*. (D. Elenius & S. Riehemann, Eds.) (1st ed., pp. 1–33). Menlo Park, California: SRI International.
- Elkin, P. L., Brown, S. H., Bauer, B. a, Husser, C. S., Carruth, W., Bergstrom, L. R., & Wahner-Roedler, D. L. (2005). A controlled trial of automated classification of negation from clinical notes. *BMC medical informatics and decision making*, 5(1), 13.
- Essex, B., & Healy, M. (1994). Evaluation of a rule base for decision making in general practice. *The British journal of general practice : the journal of the Royal College of General Practitioners*, 44(382), 211–3.
- Fagot-Largeault, A. (2010). Le concept de maladie sous-jacent aux tentatives d’informatisation du diagnostic médical. In Presses Universitaires de France - PUF (20 janvier 2010) (Ed.), *Histoire et philosophie* (1st ed., Vol. 1, pp. 89–110). Paris.
- Fenz, S. (2012). An ontology-based approach for constructing Bayesian networks. *Data & Knowledge Engineering*, 73(1), 73–88.
- Fernando Bobillo, O. (2008). *Mémoire de thèse d’université : Managing Vagueness in Ontologies*. Universidad de Granada. Retrieved from <http://digibug.ugr.es/bitstream/10481/2088/1/17701235.pdf>
- Ford, R. (2012). SWRL-IQ. *Atlassian Confluence*.
- Fragoso, G., de Coronado, S., Haber, M., Hartel, F., & Wright, L. (2004). Overview and utilization of the NCI thesaurus. *Comparative and functional genomics*, 5(8), 648–54.
- Frank, M. R., & Szekely, P. (1998). Adaptive Forms : An Interaction Technique for Entering Structured Data. *Knowledge Based Systems*, 11(1), 37–48.
- Friedman-Hill, E. (2008). Jess The rule engine for Java Platform. *Sandia National Laboratories*. Retrieved from <http://herzberg.ca.sandia.gov/>
- Gacôgne, L. (2009). Prolog, le chainage arrière. In Hermann (Ed.), *Prolog - Programmation par l’exemple* (1st ed., pp. 1–31). Paris: Hermann.
- Gangemi, A., Gomez-Perez, A., Presutti, V., & Suarez-Figueroa, C. (2008). Towards a catalog of OWL-based Ontology Design patterns. *CAEPIA*, 1(1), 1–10.
- Gangemi, A., & Presutti, V. (2009). Ontology Design Patterns. In *International Handbooks on Information Systems* (pp. 221–243).
- García-crespo, Á., Rodríguez-González, A., Mencke, M., Gómez-berbís, J. M., & Colomo-palacios, R. (2010). Expert Systems with Applications ODDIN : Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. *Expert Systems With Applications*, 37(3), 2621–2628.



## Bibliographie

- Golbeck, J., Fragoso, G., Hartel, F., Hendler, J., Oberthaler, J., & Parsia, B. (2003). The National Cancer Institute's Thésaurus and Ontology. *Web Semantics: Science, Services and Agents on the World Wide Web*, 1(1), 75–80.
- Golbreich, C., Dameron, O., Gibaud, B., & Burgun, A. (2003). Comment représenter les ontologies pour un Web Sémantique Médical ? In C. Golbreich, O. Dameron, B. Gibaud, & A. Burgun (Eds.), *Journée francophone de la toile* (pp. 1–11).
- Gomez-Perez, A. (1999). Ontological Engineering: A state of the art. *Expert Update*, 1(1), 33–43.
- Gordon, T. F., Governatori, G., & Rotolo, A. (2009). Rules and Norms : Requirements for Rule Interchange Languages in the Legal Domain. In G. Governatori, J. Hall, & A. Paschke (Eds.), *The International RuleML Symposium on Rule Interchange and Applications* (pp. 282–296). Las Vegas, Nevada, USA: Springer Berlin Heidelberg.
- Grigorova, D., & Nikolov, N. (2008). Knowledge Representation and Reasoning in Natural Language Processing Systems. *International Scientific Conference Computer Science*, 74(7), 641–646.
- Grosbois, B., Decaux, O., Guenet, L., Goasguen, J., & Jegou, P. (2009). Nosology and management of monoclonal gammopathy. *Bulletin de l'académie nationale de médecine*, 193(5), 1069–1085.
- Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Creation Diffusion Utilization*, 5(2), 199–220.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5-6), 907–928.
- Grundstein, M., De Bonnières, P., & Para, S. (1988a). La représentation des connaissances, la reproduction des raisonnements, la simulation des tâches. In M. Grundstein, P. De Bonnières, & S. Para (Eds.), *Les systèmes à base de connaissances* (1st ed., pp. 76–101). Paris - La défense: Afnor Gestion.
- Grundstein, M., De Bonnières, P., & Para, S. (1988b). La découverte par comparaison. In M. Grundstein, P. De Bonnières, & S. Para (Eds.), *Les systèmes à base de connaissances* (1st ed., pp. 32–55). Paris - La défense: Afnor Gestion.
- Haarslev, V., & Ralf, M. (2000). Consistency Testing: The RACE Experience. In *Proceedings, Automated Reasoning with Analytic tableau and related methods* (pp. 57–61).
- Haav, H., & Lubi, T. (2001). A Survey of Concept-based Information Retrieval Tools on the Web. In H. Haav & T. Lubi (Eds.), *5th East-European Conference, ADBIS 2001* (pp. 29–41). Vilnius: Vilnius “Technika.”
- Harrison, N. B., Avgeriou, P., & Zdun, U. (2007). Using Patterns to Capture Architectural Decisions. *IEEE Software*, 24(4), 38–45.

## Bibliographie

- Haute Autorité de Santé publique. (2008). *Prise en charge thérapeutique des lésions méniscales et des lésions isolées du ligament croisé antérieur du genou chez l'adulte. Recherche* (pp. 1–105). Saint-Denis La Plaine. Retrieved from [http://www.has-sante.fr/portail/upload/docs/application/pdf/2008-07/lesions\\_meniscales\\_et\\_du\\_ligament\\_croise\\_anterieur\\_-\\_argumentaire.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/2008-07/lesions_meniscales_et_du_ligament_croise_anterieur_-_argumentaire.pdf)
- Hayes, P., Carroll, J., Welty, C., Uschold, M., Vatan, B., Manola, F., ... Lawrence, J. (2006). Defining N-ary Relations on the Semantic Web. *W3C Working Group Note*. Retrieved from <http://www.w3.org/TR/swbp-n-aryRelations/>
- Health Service Ombudsman for England. (1998). *Errors in the care and treatment of a young woman with diabetes* (pp. 1–32).
- Heckerman, D., Horvitz, E., & Nathwani, B. (1992). Toward normative expert systems: Part I. The Pathfinder project. *Methods of information in medicine*, 31(2), 90–105.
- Henson, C., Thirunarayan, K., Sheth, A., & Hitzler, P. (2011). Representation of Parsimonious Covering Theory in OWL-DL. In M. Dumontier (Ed.), *Proceedings of the 8th International Workshop on OWL: Experiences and Directions (OWLED 2011)* (Vol. 25, pp. 1–5). San Francisco, CA, USA.
- Horn, A. (1951). On Sentences Which are True of Direct Unions of Algebras. *The Journal of Symbolic Logic*, 16(1), 14–21.
- Horridge, M., & Bechhofer, S. (2009). The OWL API: A Java API for Working with OWL 2 Ontologies. Manchester. Retrieved from [http://webont.org/owled/2009/papers/owled2009\\_submission\\_29.pdf](http://webont.org/owled/2009/papers/owled2009_submission_29.pdf)
- Horridge, M., Knublauch, H., Rector, A., Stevens, R., & Wroe, C. (2004). *Building An OWL Ontology*. (The University of Manchester and Stanford University, Ed.) *A Practical Guide To Building OWL Ontologies Using The Protégé OWL Plugin and CO-ODE Tools - Edition 1.0* (Edition 1., pp. 16–81). Manchester.
- Horrocks, I., Patel-Schneider, P., Boley, H., Macgregor, S. T., Benjamin, G., & Mike, D. (2004). SWRL: A Semantic Web Rule Language Combining OWL and RuleML. *W3C Member*.
- Horrocks, I., & Patel-schneider, P. F. (2004). A Proposal for an OWL Rules Language. In I. Horrocks & P. F. Patel-schneider (Eds.), *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)* (pp. 723–731). ACM.
- Horrocks, J., McCann, P., Staniland, J., Leaper, D., & De Dombal, F. (1972). Computer-aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases. *British medical journal*, 2(5804), 5–9.
- Jakob, C., Sterz, J., Zavrski, I., Heider, U., Kleeberg, L., Fleissner, C., ... Sezer, O. (2006). Angiogenesis in multiple myeloma. *European journal of cancer*, 42(11), 1581–90.
- Javanmardi, S., Amini, M., & Jalili, R. (2006). An Access Control Model for Protecting Semantic Web Resources. In *Web policy workshop Ahens, GA, USA* (pp. 1–15).



## Bibliographie

- Josephson, J., Josephson, S., & Tanner, M. (1996). Conceptual analysis of abduction. In J. R. Josephson & S. G. Josephson (Eds.), *Abductive Inference - Computation, Philosophy, Technology* (1st ed., pp. 1–29). Cambridge: Cambridge University Press.
- Jovic, A., Prcela, M., & Gamberger, D. (2007). Ontologies in Medical Knowledge Representation. In *2007 29th International Conference on Information Technology Interfaces* (pp. 535–540). IEEE.
- Kalenderian, E., Ramoni, R. L., White, J. M., Schoonheim-Klein, M. E., Stark, P. C., Kimmes, N. S., ... Walji, M. F. (2011). The Development of a Dental Diagnostic Terminology. *J Dent Educ*, 75(1), 68–76.
- Kassel, G. (2002). OntoSpec : une méthode de spécification semi-informelle d'ontologies. In *Actes des journées francophones d'Ingénierie des Connaissances (IC'2002)*, (pp. 75–87). Paris: IC'2002.
- Kayaalp, M. (2005). Why do we need probabilistic approaches to ontologies and the associated data ? In *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* (p. 1005).
- Kiryakov, A., Ognyanov, D., & Manov, D. (2005). OWLIM – a Pragmatic Semantic Repository for OWL. In *WISE Workshops, Springer (2005)* (pp. 182–192).
- Kulikowski, A. C., & Weiss, S. M. (1982). Representation of Expert Knowledge for Consultation: The CASNET and EXPERT Projects. In B. Westview Press (Ed.), *Artificial Intelligence in Medicine* (1st ed., p. 1). Colorado: Szolovits Peter.
- Kyle, R., Francis, B., & Vincent, R. (2011). Management of monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). *Oncology*, 25(7), 578–586.
- Laskey, Kathryn B, & Costa, P. C. G. (2005). Of Starships and Klingons : Bayesian Logic for the 23 rd Century. In Kathryn B Laskey & P. C. G. Costa (Eds.), *Proceedings of the Twenty-first Conference (UAI 2005)* (pp. 1–8). Edinburgh, Scotland: AUAI Press.
- Laskey, Kathryn Blackmond, & Haberlin, R. (2011). PR-OWL 2 Case Study : A Maritime Domain Probabilistic Ontology. In Kathryn Blackmond Laskey & R. Haberlin (Eds.), *Proceedings of the Sixth International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS 2011)* (pp. 1–8). STIDS 2011.
- Laurent, D. (2007). Introduction aux patterns de conception. In ENI (Ed.), *Design Patterns - Les 23 modèles de conception: descriptions et solutions illustrées en UML 2 et Java*. (1st ed., pp. 17–26). Saint Herblain: Joelle Musset.
- Le Moing, C. (2009). *Mémoire de thèse d'exercice : Système informatisé d'aide au diagnostic en urgences odontologiques*. Université de Rennes 1. Retrieved from [www.sudoc.fr/135996198](http://www.sudoc.fr/135996198)

## Bibliographie

- Lemaire, J. B., Schaefer, J. P., Martin, L. A., Faris, P., Ainslie, M. D., & Hull, R. D. (1999). Effectiveness of the quick medical reference as a diagnostic tool. *CMAJ*, 161(6), 725–728.
- Lin, C., Luciani, A., Belhadj, K., Deux, J.-F., Kuhnowski, F., Maatouk, M., ... Rahmouni, A. (2010). Multiple Myeloma Treatment Response Assessment with Whole-Body Dynamic Contrastenhanced MR Imaging. *Radiology*, 254(2), 521–531.
- Lucas, P. J. F. (2001). Certainty-factor-like structures in Bayesian belief networks. *Knowledge-Based Systems*, 14(7), 327–335.
- Maedche, A., & Staab, S. (2001). Ontology Learning for the semantic web. *IEEE Intelligent Systems*, 16(2), 72–79.
- Malviya, N., Mishra, N., & Sahu, S. (2011). Developing University Ontology using protégé OWL Tool : Process and Reasoning. *International journal of scientific and engineering research*, 2(9), 1–8.
- Masquelet, A. C. (2006a). Le raisonnement diagnostique. In P. U. de F. PUF (Ed.), *Le raisonnement médical* (1st ed., pp. 17–43). Paris: PUF, Presses Universitaires de France.
- Masquelet, A. C. (2006b). Les modes fondamentaux d'inférence. In P. U. de F.- PUF (Ed.), *Le raisonnement médical* (1st ed., pp. 75–93). Paris: PUF, Presses Universitaires de France.
- Matheron, J. P. (2005). Le modèle conceptuel des données. In J. patrick Matheron (Ed.), *Comprendre MERISE - Outils conceptuels et organisationnels* (1st ed., pp. 39–78). Paris: Eyrolles.
- Mccormick, R. (1997). Conceptual and Procedural Knowledge. *International Journal of Technology and design Education*, 7(1), 141–159.
- McGee, S. (2002). Simplifying likelihood ratios. *Journal of general internal medicine*, 17(8), 646–9.
- McGuinness, D. L., & Harmelen, F. van. (2004). OWL Web Ontology Language Overview. *W3C Working Group Note*. Retrieved from <http://www.w3.org/TR/owl-features/>
- Mendonça, E. (2004). Clinical decision support systems: perspectives in dentistry. *Journal of dental education*, 68(6), 589–97.
- Miller, R. A. (2009). Computer-assisted diagnostic decision support: history, challenges, and possible paths forward. *Advances in health sciences education : theory and practice*, 14(1), 89–106.
- Minsky, M. (1975). Minsky'S Frame System Theory. In *TINLAP '75 Proceedings of the 1975 workshop on Theoretical issues in natural language processing* (pp. 104–116).
- Mizoguchi, R., & Ikeda, M. (1998). Towards Ontology Engineering. In *Technical Report* (pp. 1–10).

## Bibliographie

- Moehler, T. ., Ho, A. ., Goldschmidt, H., & Barlogie, B. (2003). Angiogenesis in hematologic malignancies. *Critical Reviews in Oncology/Hematology*, 45(3), 227–244.
- Motik, B., & Sattler, U. (2006). A Comparison of Reasoning Techniques for Querying Large Description Logic ABoxes. In *the 13th international conference on Logic for Programming, Artificial Intelligence, and Reasoning* (pp. 227–241).
- Motik, B., Sattler, U., & Studer, R. (2005). Query Answering for OWL-DL with Rules. *Journal of Web Semantics : Science, Services and Agents on the World Wide Web*, 1(1), 41–60.
- Munshi, N. (2008). Investigative tools for diagnosis and management. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program*, 2008(1), 298–305.
- Murphy, G. C., Friedman, C. P., Elstein, a S., Wolf, F. M., Miller, T., & Miller, J. G. (1996). The influence of a decision support system on the differential diagnosis of medical practitioners at three levels of training. *Proceedings : a conference of the American Medical Informatics Association / ... AMIA Annual Fall Symposium. AMIA Fall Symposium*, 1(1), 219–23.
- Musen, M. A. (1992). Dimensions of knowledge sharing and reuse. *Computers and Biomedical Research*, 25(5), 435–467.
- Naïm, P., Wuillemin, P., Leray, P., Pourret, O., & Becker, A. (2007). Introduction aux réseaux bayésiens. In Eyrolles (Ed.), *Réseaux bayesiens* (pp. 20–22). Paris: Eyrolles.
- Nakache, D., & Metais, E. (2005). Evaluation: Nouvelle approche avec juges. *INFORSID*, 1(1), 1–15.
- Neapolitan, R. (2003). Introduction to Bayesian Networks. In Prentice Hall Series in Artificial Intelligence (Ed.), *Learning Bayesian Networks* (1st ed., pp. 1–10). Chicago, Illinois: Prentice Hall.
- Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. R. (1991). Enabling Technology for Knowledge Sharing. *Artificial Intelligence Magazine*, 12(3), 36–56.
- Negnevitsky, M. (2002a). Frame based expert systems. In Addison - Wesley (Ed.), *Artificial intelligence - A guide to intelligent systems* (1st ed., pp. 129–162). Edinburgh Gate: Addison - Wesley.
- Negnevitsky, M. (2002b). Uncertainty management in rule based expert systems. In Addison Wesley (Ed.), *Artificial intelligence - A guide to intelligent systems* (2nd ed., pp. 54–86). Edinburgh Gate: Addison Wesley.
- Negnevitsky, M. (2002c). Rule based expert systems. In Addyson - Wesley (Ed.), *Artificial intelligence - A guide to intelligent systems* (2nd ed., pp. 25–54). Edinburgh Gate: Wesley, Addison.

## Bibliographie

- Newell, A. (1973). *Productions systems : models of control structures* (Vol. 1, pp. 1–68). Canergie Mellon.
- Noy, N. F., de Coronado, S., Solbrig, H., Fragoso, G., Hartel, W. F., & Musen, M. A. (2008). Representing the NCI Thesaurus in OWL DL: Modeling tools help modeling languages. *Applied ontology*, 3(3), 173–190.
- O'Connor, M., & Das, A. (2009). SQWRL : a Query Language for OWL. In M. O'Connor & A. Das (Eds.), *6th International Workshop on OWL: Experiences and Directions (OWLED 2009)* (pp. 1–8). Amsterdam, Netherlands: Vrije Universiteit.
- O'Connor, M., Nyulas, C., Shankar, R., Das, A., & Musen, M. (2008). The SWRLAPI : A Development Environment for Working with SWRL Rules The SWRLAPI : Core Features. In Technical Report (Ed.), *OWL: Experiences and Directions (OWLED), Fifth International Workshop, held with 7th International Semantic Web Conference* (pp. 1–4). Karlsruhe, Germany.
- Oracle. (2011). MySQL. *Oracle Corporation*. Retrieved from <http://www.oracle.com/fr/products/mysql/index.html>
- Page, M. (1990). *Mémoire de thèse d'université : Système expert à base de connaissances profondes: Application à un poste de travail "intelligent" pour le comptable*. Institut National Polytechnique de Grenoble. Retrieved from <http://hal.inria.fr/tel-00338750/>
- Patel, V., Arocha, J., & ZHANG, J. (2004). Thinking and Reasoning in Medicine Citation : What is Medical Reasoning? Early Research on Medical Problem Solving and Reasoning, 1–34.
- Patel-Schneider, P. F. (2009). OWL 2 Web Ontology Language New Features and Rationale. *W3C Working Group Note*. Retrieved from [www.w3.org/TR/2008/WD-owl2-new-features-20081202](http://www.w3.org/TR/2008/WD-owl2-new-features-20081202)
- Peelen, L., Klein, M. C. A., Schlobach, S., Keizer, N. F. De, & Peek, N. (2007). Analyzing Differences in Operational Disease Definitions Using Ontological Modeling. In L. Peelen, M. C. A. Klein, S. Schlobach, N. F. De Keizer, & N. Peek (Eds.), *AIME '07 Proceedings of the 11th conference on Artificial Intelligence in Medicine* (pp. 297–302). Springer-Verlag Berlin, Heidelberg.
- Peter, I. H., Patel-Schneider, F., Boley, H., Tabet, S., Grosof, B., & Dean, M. (2004). SWRL Section 8. Built-Ins. *DARPA DAML Program*. Retrieved from <http://www.daml.org/rules/proposal/builtins>
- Pottier, P., & Planchon, B. (2011). Description of the mental processes occurring during clinical reasoning. *La Revue de médecine interne / fondée ... par la Société nationale française de médecine interne*, 32(6), 383–90.
- Pradeep, M., & Prakash, N. (2001). Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents: *JAMIA*, 8(6), 598–609.

## Bibliographie

- Prcela, M., Gamberger, D., & Jovic, A. (2008). Semantic Web Ontology Utilization for Heart Failure Expert System Design. *Studies in Health Technology and Informatics*, 136(1), 851–856.
- ProtegeWiki. (2012). Protege 3.4.8. *Stanford*. Retrieved from [http://protegewiki.stanford.edu/wiki/Protege\\_3.4.8\\_Release\\_Notes](http://protegewiki.stanford.edu/wiki/Protege_3.4.8_Release_Notes)
- Rajkumar, V. S. (2011). Multiple myeloma: 2011 update on diagnosis, risk-stratification, and management. *American journal of hematology*, 86(1), 57–65.
- RDF Working Group. (2004). Resource Description Framework (RDF). *W3C*. Retrieved from <http://www.w3.org/RDF/>
- Rector, A., & Schreiber, G. (2005). Qualified cardinality restrictions (QCRs): Constraining the number of values of a particular type for a property. *W3C*. Retrieved from <http://www.w3.org/2001/sw/BestPractices/OEP/QCR/>
- Reggia, J. A., & Peng, Y. (1987). Modeling Diagnostic Reasoning: A Summary of Parsimonious Covering Theory. *Computer Methods and Programs Biomedicine*, 25(1), 125–134.
- Rodríguez-González, A., & Alor-Hernández, G. (2013). An approach for solving multi-level diagnosis in high sensitivity medical diagnosis systems through the application of semantic technologies. *Computers in biology and medicine*, 43(1), 51–62.
- Rodríguez-González, A., García-Crespo, Á., Colomo-Palacios, R., Gayo, J. E. L., Gómez-Berbis, J. M., & Alor-Hernández, G. (2011). Automated Diagnosis Through Ontologies and Logical Descriptions. *International Journal of Decision Support System Technology*, 3(1), 21–39.
- Rodriguez-Gonzalez, A., Hernandez-Chan, G., Colomo-Palacios, R., Gomez-Berbis, J. M., Garcia-Crespo, A., Alor-Hernandez, G., & Valencia-Garcia, R. (2010). Towards an Ontology to Support Semantics Enabled Diagnostic Decision Support Systems. *Semantic Web and Healthcare*, 7(3), 234–245.
- Rodríguez-González, A., Labra-Gayo, J. E., Colomo-Palacios, R., Mayer, M. a, Gómez-Berbis, J. M., & García-Crespo, A. (2012). SeDeLo: using semantics and description logics to support aided clinical diagnosis. *Journal of medical systems*, 36(4), 2471–81.
- Rodríguez-González, A., Torres-Niño, J., Mayer, M., Alor-Hernandez, G., & Wilkinson, M. D. (2012). Analysis of a Multilevel Diagnosis Decision Support System and Its Implications: A Case Study. *Computational and Mathematical Methods in Medicine*, 2012(1), 1–9.
- Rosse, C., & Mejino, J. L. V. (2003). A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of biomedical informatics*, 36(6), 478–500.
- Scarlett, P. Y., Cooke, M. W., Clarke, D., Bates, C., & Chan May, A. (1986). Computer aided diagnosis of acute abdominal pain at Middlesbrough General Hospital. *Annals of the royal college of surgeons of england*, 68(4), 179–181.



## Bibliographie

- Scheuermann, R. H., Ceusters, W., & Smith, B. (2009). Toward an Ontological Treatment of Disease and Diagnosis Department of Pathology and Division of Biomedical Informatics , University of Texas. In *AMIA 2009 summit on translational bioinformatics* (pp. 116–120).
- Seidenberg, J., & Rector, A. (2006). Representing Transitive Propagation in OWL. In J. Seidenberg & A. Rector (Eds.), *Proc. of the 25th Int. Conf. on Conceptual Modeling (ER 2006)* (pp. 255–266). Springer.
- Seka, L. ., Pouliquen, B., & Le Beux, P. (1996). Indexation et interrogation automatique de textes médicaux: application à la base A.D.M. In P. Moukeli (Ed.), *CARI'96 : actes du 3ème colloque africain sur la recherche en informatique = CARI'96 : proceedings of the 3rd African conference on research in computer science* (pp. 247–261). Paris.
- Sensoy, M., Fokoue, A., Srivatsa, M., & Pan, J. Z. (2011). Probabilistic Abox Abduction in Description Logics. In *SSWS* (pp. 106–121).
- Seon Gyu, P., & Hong-Gee, K. (2006). Dental Decision Making on Missing Tooth Represented in an Ontology and Rules. *Lecture Notes in Computer Science*, 4185, 322–328.
- Seon Gyu, P., Hong-Gee, K., & Myeng-Ki, K. (2007). Tooth Positional Ontology Represented in OWL. In *Medinfo* (pp. 338–9).
- Seroussi, B., Le Beux, P., & Venot, A. (2013). L'aide au diagnostic médical. In Springer Verlag France (Ed.), *Informatique médicale, e santé. Fondements et applications* (1st ed., pp. 147–175). Paris: Springer. Retrieved from <http://link.springer.com/book/10.1007%2F978-2-8178-0338-8>
- Shearer, R., Motik, B., & Horrocks, I. (2008). HermiT : A Highly-Efficient OWL Reasoner. In *OWLED* (pp. 1–10).
- Shuttleworth, M. (2013). Operationalization. *Explorable.com*. Retrieved March 14, 2013, from <http://explorable.com/operationalization.html>
- Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical OWL-DL reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2), 51–53.
- Sirohi, B., & Powles, R. (2004). Multiple myeloma. *Lancet*, 363(9412), 875–87.
- Smets, P. (2005). Belief Functions : the Disjunctive Rule of Combination and the Generalized Bayesian. *International Journal of Approximate Reasoning*, 9(1), 1–32.
- Soualmia, L. F., Golbreich, C., & Darmoni, S. J. (2004). Representing the MeSH in OWL : Towards a Semi-Automatic Migration. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*.

## Bibliographie

- Stenzhorn, H., Gabriele, W., Mathias, B., Fatima, S., Vangelis, K., Manolis, T., ... Norbert, G. (2010). The ObTiMA system - ontology-based managing of clinical trials. *Studies in Health Technology and Informatics*, 160(2), 1090 – 1094.
- Stevens, R., Egaña Aranguren, M., Wolstencroft, K., Sattler, U., Drummond, N., Horridge, M., & Rector, A. (2007). Using OWL to Model Biological Knowledge. *International Journal of Human-Computer Studies*, 65(7), 583–594.
- Straccia, U. (1998). A Fuzzy Description Logic. In *Proceedings of AAAI-98, 15th National Conference on Artificial Intelligence* (pp. 1–6). Madison, Wisconsin.
- Studer, R., Benjamins, R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1-2), 161–197.
- The Apache Software Foundation. (2008). JenaRules. *The Apache Software Foundation*.
- Tsarkov, D., & Horrocks, I. (2006). FaCT ++ Description Logic Reasoner : System Description. In *Automated Reasoning (IJCAR 2006)* (pp. 292–297).
- Turban, A., & Liang. (2005). Knowledge acquisition, representation, and reasoning. In P. Hall (Ed.), *The scope and types of knowledge* (7th ed., pp. 174–253).
- Uschold, M., & King, M. (1995). Towards a methodology for building ontologies. In M. Uschold & M. King (Eds.), *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI'95* (pp. 1–15). Edinburgh, UK: Uschold, Mike King, Martin.
- Van Melle, W. (1984). The Structure of the MYCIN System. In E. H. Shortliffe & B. G. Buchanan (Eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (1st ed., Vol. 322, pp. 67–77). Menlo Park, California.
- Vertue, F. M., & Haig, D. B. (2008). An abductive perspective on clinical reasoning and case formulation. *Journal of Clinical Psychology*, 64(9), 1046–1068.
- Vikram, K., & Karjodkar, F. R. (2009). Decision support systems in dental decision making: an introduction. *The journal of evidence-based dental practice*, 9(2), 73–6. doi:10.1016/j.jebdp.2009.03.003
- Walton, D. (2004). Abductive inference in medical diagnosis. In D. Walton (Ed.), *Abductive reasoning* (1st ed., pp. 143–157). Alabama: University Alabama Press.
- Wang, H. H., Tu, S., Noy, N., Rector, A., Musen, M., Redmond, T., ... Seidenberg, J. (2006). Frames and OWL Side by Side. In *9th International Protégé Conference (2006)* (pp. 1–4).
- Warner, H. R., Haug, P., Bouhaddou, O., Lincoln, M., Wamer, H., Sorenson, D., ... Fan, C. (1988). ILIAD as an expert consultatnt to teach Differential Diagnosis. In H. R. Warner, P. Haug, O. Bouhaddou, M. Lincoln, H. Wamer, D. Sorenson, ... C. Fan (Eds.), *Annual Symposium on Computer Application in Medical Care* (pp. 371–376). Salt Lake City.



## Bibliographie

- Watanaboonyongcharoen, P., Nakorn, T. N., Rojnuckarin, P., Lawasut, P., & Intragumtornchai, T. (2012). Prevalence of monoclonal gammopathy of undetermined significance in Thailand. *International Journal of Hematology*, 95(2), 176–81.
- Weiss, S. M., Kulikowski, C. A., & Safir, A. (1977). A Model based consultation system for the long term management of glaucoma. *IJCAI*, 3(5), 826–832.
- Welty, C., & Guarino, N. (2001). Supporting ontological analysis of taxonomic relationships. *Data & Knowledge Engineering*, 39(1), 51–74.
- Winograd, T. (1975). Frame Representations and the Declarative-Procedural Controversy. In D. Bobrow & A. Collins (Eds.), *Representation and Understanding: Studies in Cognitive Science* (pp. 185–210). New York 1975.
- Wong, J. M.-L., Khan, T., Jayadev, C. S., Khan, W., & Johnstone, D. (2012). Anterior cruciate ligament rupture and osteoarthritis progression. *The open orthopaedics journal*, 6, 295–300.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), 338–353.
- Zalila, Z., Cuquemelle, J., Penet, C., Chikh, A., Lorentz, B., Deschamps, D., & Assemat, C. (2008). Logique floue et systèmes d ' inférence floue Introduction et propriétés. Retrieved from <http://www.xtractis.fr/pdfDoc/White Paper FR - Introduction a la theorie du flou.pdf>
- Zhang, J., Patel, V. L., Johnson, T. R., & Shortliffe, E. H. (2004). A cognitive taxonomy of medical errors. *Journal of biomedical informatics*, 37(3), 193–204.
- Zhang, J., Patel, V., Turley, J., & Johnson, T. (2005). Health Informatics and Medical Errors. *Business Briefing North American Pharmacotherapy*, 1(1), 1–4.
- Zins, C. (2006). Conceptual Approaches for Defining Data, Information, and Knowledge. *Journal of the american society for information science and technology*, 58(4), 479–493.

# Annexes

## Annexe 1: Classe « Java » de chargement de l'ontologie OWL pour l'instanciation automatique.

```

FileOWL.java  PopulateOwlFile.java  DbActionOnto.java  DentalDiagnose.java

package com.main;

import com.owl.FileOWL;
import com.owl.FileOWLManager;

public class PopulateOwlFile {
    public static void print(String texte) {
        System.out.println(texte);
    }

    /**
     * @param args
     */
    public static void main(String[] args) {
        if (args.length == 3) {
            try {
                FileOWL file = FileOWLManager.getInstance().loadFromFile(args[0]);
                if (new Integer(args[2]).intValue() == 0) {
                    print(" Code Error 0: Le nombre d'instance à créer doit être supérieur à zéro. ");
                }
                else {
                    // stockage des classes définies dans une base de données
                    // FileOWLManager.getInstance().storeDefinedClassToKb(file);
                    // stockage des classes définies dans une base de données
                    // créer les instances
                    FileOWLManager.getInstance().createInstance(file, new Integer(args[2]).intValue());
                    // Assertionnement;
                    FileOWLManager.getInstance().AddAssertionObjectProperties(file);
                    // Assertionnement
                    // sauvegarder
                    FileOWLManager.getInstance().saveToFile(file, args[1]);
                }
            } catch (Exception e) {
                print(e.getMessage());
            }
        }
        else {
            print("Code Error 1: Nombre d'argument incorrect.");
        }
    }
}

```

## Annexe 2: Classe « Java » de chargement des ensembles de triplets dans la base de données relationnelle.

```

FileOWL.java  PopulateOwlFile.java  DbActionOnto.java  DentalDiagnose.java
public static void main(String[] args) throws SQWRLException {
    // TODO Auto-generated method stub
    final String ONTOLOGY_URL = "file:///C:/Doc/MyelomaOntology07122011.owl";
    DbActionOnto dbonto1 = new DbActionOnto();
    try {
        dbonto1.doConnection("com.mysql.jdbc.Driver", "jdbc:mysql://localhost/dbdiagnose1?" + "user=root&password=");
    } catch (ClassNotFoundException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    } catch (SQLException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
    System.out.println(" file : " + ONTOLOGY_URL);
    SQWRLQueryEngine queryEngine;
    SWRLRuleEngine rule = null;
    OWLModel owlModel = null;
    try {
        owlModel = ProtegeOWL.createJenaOWLModelFromURI(ONTOLOGY_URL);
    } catch (OntologyLoadException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
    try {
        rule = SWRLRuleEngineFactory.create(owlModel);
        rule.importSWRLRulesAndOWLKnowledge();
        rule.run();
    } catch (SWRLRuleEngineException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }
    queryEngine = SQWRLQueryEngineFactory.create(owlModel);
    // queryEngine.createSQWRLQuery("Finding_Has_Diagnosis_Name", "Finding(?finding) ^ Disease_May_Have_Finding(?disease, ?fir
    queryEngine.runSQWRLQueries();
    SQWRLResult Finding_Absence_Exclude_Diagnosis = queryEngine.getSQWRLResult("Query_Finding_Absence_Excludes_Diagnosis");
    SQWRLResult Finding_Absence_Has_Diagnosis = queryEngine.getSQWRLResult("Query_Finding_Absence_Has_Diagnosis");
    SQWRLResult Finding_Exclude_Diagnosis = queryEngine.getSQWRLResult("Query_Finding_Excludes_Diagnosis");
    SQWRLResult Finding_Has_Diagnosis = queryEngine.getSQWRLResult("Query_Finding_Has_Diagnosis");
    Finding_Absence_Exclude_Diagnosis.reset();
    Finding_Absence_Has_Diagnosis.reset();
    Finding_Exclude_Diagnosis.reset();
    Finding_Has_Diagnosis.reset();
    int NbreLigne = new Integer(Finding_Absence_Exclude_Diagnosis.getNumberOfRows() + Finding_Absence_Has_Diagnosis.getNumberOfRows());
}

```

### **Annexe 3: Classe « Java » de chargement des restrictions de l'ontologie en mémoire.**

```
package com.onto.auth.beans.ontology;

import org.apache.struts.action.ActionForm;

public class OntologyRestrictionBean extends ActionForm implements Comparable {

    private static final long serialVersionUID = 1L;

    private Integer conceptId;

    private Float sensibilite;

    private Float specificite;

    private Integer roleId;

    private String role;

    private Integer diseaseId;

    private String disease;

    private Integer findingId;

    private String finding;

    private String ontology;

    private Float youdenIndication;

    private Float rvp;

    private Float rvn;

    private Integer Concept_Restriction_seuil_Id;

    private String Concept_Restriction_symbole;

    private String Concept_Restriction_Unite;

    private String Concept_Restriction_valeur;

    private Integer Type_Zone;

    public OntologyRestrictionBean(Integer conceptId, Float sensibilite, Float specificite,
Integer roleId, String role, Integer diseaseId, String disease, Integer findingId,
```

String finding, String ontology, Float youdenIndication, Float rvp, Float  
rvn, Integer Concept\_Restriction\_seuil\_Id, String Concept\_Restriction\_symbole,

String Concept\_Restriction\_Unite, String Concept\_Restriction\_valeur, Integer  
Type\_Zone) {

```

    super();

    this.conceptId = conceptId;

    this.sensibilite = sensibilite;

    this.specificite = specificite;

    this.roleId = roleId;

    this.role = role;

    this.diseaseId = diseaseId;

    this.disease = disease;

    this.findingId = findingId;

    this.finding = finding;

    this.ontology = ontology;

    this.youdenIndication = youdenIndication;

    this.rvp = rvp;

    this.rvn = rvn;

    this.Concept_Restriction_seuil_Id = Concept_Restriction_seuil_Id;

    this.Concept_Restriction_symbole = Concept_Restriction_symbole;

    this.Concept_Restriction_Unite = Concept_Restriction_Unite;

    this.Concept_Restriction_valeur = Concept_Restriction_valeur;

    this.Type_Zone = Type_Zone;

```

}

@Override

```

public String toString() {

    return "OntologyRestrictionBean [conceptId=" + conceptId

        + ", sensibilite=" + sensibilite + ", specificite="

```

```

+ specificite + ", roleId=" + roleId + ", role=" + role
+ ", diseaseId=" + diseaseId + ", disease=" + disease
+ ", findingId=" + findingId + ", finding=" + finding
+ ", ontology=" + ontology + ", youdenIndication="
+ youdenIndication + ", rvp=" + rvp + ", rvn=" + rvn
+ ", Concept_Restriction_seuil_Id="
+ Concept_Restriction_seuil_Id
+ ", Concept_Restriction_symbole="
+ Concept_Restriction_symbole + ",
Concept_Restriction_Unite="
+ Concept_Restriction_Unite + ", Concept_Restriction_valeur="
+ Concept_Restriction_valeur + ", Type_Zone=" + Type_Zone
+ "]" ;

}

public Integer getConcept_Restriction_seuil_Id() {
    return Concept_Restriction_seuil_Id;
}

public void setConcept_Restriction_seuil_Id(Integer concept_Restriction_seuil_Id) {
    Concept_Restriction_seuil_Id = concept_Restriction_seuil_Id;
}

public String getConcept_Restriction_symbole() {
    return Concept_Restriction_symbole;
}

public void setConcept_Restriction_symbole(String concept_Restriction_symbole) {
    Concept_Restriction_symbole = concept_Restriction_symbole;
}

public String getConcept_Restriction_Unite() {
    return Concept_Restriction_Unite;
}

```

```
}

public void setConcept_Restriction_Unite(String concept_Restriction_Unite) {

    Concept_Restriction_Unite = concept_Restriction_Unite;

}

public String getConcept_Restriction_valeur() {

    return Concept_Restriction_valeur;

}

public void setConcept_Restriction_valeur(String concept_Restriction_valeur) {

    Concept_Restriction_valeur = concept_Restriction_valeur;

}

public Integer getType_Zone() {

    return Type_Zone;

}

public void setType_Zone(Integer type_Zone) {

    Type_Zone = type_Zone;

}

public Integer getRoleId() {

    return roleId;

}

public void setRoleId(int roleId) {

    this.roleId = roleId;

}

public Integer getConceptId() {

    return conceptId;

}

public void setConceptId(int conceptId) {

    this.conceptId = conceptId;

}
```



```
}

public Float getSensibilite() {
    return sensibilite;
}

public void setSensibilite(Float sensibilite) {
    this.sensibilite = sensibilite;
}

public Float getSpecificite() {
    return specificite;
}

public void setSpecificite(Float specificite) {
    this.specificite = specificite;
}

public String getRole() {
    return role;
}

public void setRole(String role) {
    this.role = role;
}

public Integer getDiseaseId() {
    return diseaseId;
}

public void setDiseaseId(int diseaseId) {
    this.diseaseId = diseaseId;
}

public String getDisease() {
    return disease;
}
```

```
    }

    public void setDisease(String disease) {
        this.disease = disease;
    }

    public Integer getFindingId() {
        return findingId;
    }

    public void setFindingId(int findingId) {
        this.findingId = findingId;
    }

    public String getFinding() {
        return finding;
    }

    public void setFinding(String finding) {
        this.finding = finding;
    }

    public String getOntology() {
        return ontology;
    }

    public void setOntology(String ontology) {
        this.ontology = ontology;
    }

    public Float getYoudenIndication() {
        return youdenIndication;
    }

    public void setYoudenIndication(Float youdenIndication) {
        this.youdenIndication = youdenIndication;
    }
}
```

```

    }

    public Float getRvp() {
        return rvp;
    }

    public void setRvp(Float rvp) {
        this.rvp = rvp;
    }

    public Float getRvn() {
        return rvn;
    }

    public void setRvn(Float rvn) {
        this.rvn = rvn;
    }

    @Override
    public int compareTo(Object arg0) {
        OntologyRestrictionBean oBean = (OntologyRestrictionBean) arg0;

        // TODO Auto-generated method stub

        return
(this.disease+"#" +this.conceptId).compareTo(oBean.getDisease()+"#" +oBean.conceptId);
    }

    @Override
    public int hashCode() {
        final Integer prime = 31;

        int result = 1;

        result = prime * result + ((conceptId == null) ? 0 : conceptId.hashCode());

        return result;
    }

```

```
@Override
public boolean equals(Object obj) {
    if (this == obj)
        return true;
    if (obj == null)
        return false;
    if (getClass() != obj.getClass())
        return false;
    OntologyRestrictionBean other = (OntologyRestrictionBean) obj;
    if (conceptId == null) {
        if (other.conceptId != null)
            return false;
    } else if (!conceptId.equals(other.conceptId))
        return false;
    return true;
}
}
```

## Liste des publications et communications

### Publications dans des revues internationales

Donfack Guefack, V., Bertaud-Gounot, V., Duvauferrier, R., Bourdé, A., Morelli, J., & Lasbleiz, J. (2012). **Ontology Driven Decision Support Systems for Medical Diagnosis- An interactive form for consultation in patients with plasma cell disease.** *Studies in Health Technology and Informatics*, 180(1), 108–112.

Bertaud-Gounot, V., Donfack Guefack, V., Lasbleiz, J., Bourdé, A., & Duvauferrier, R. (2011). **Creating an ontology driven rules base for an expert system for medical diagnosis.** *Studies in Health Technology and Informatics*, 169, 714–718.

### Communications internationales

Bertaud-Gounot, V., Donfack Guefack, V., Brillet, E., & Duvauferrier, R. (2012). **Les technologies du web sémantique pour un renouveau des systèmes experts en médecine. Principes, problèmes et propositions à partir de l'exemple du myélome dans le NCI-T.** Quatorzièmes Journées francophones d'informatique médicale (JFIM) Tunis-Tunisie, 30 mars-2 Avril 2011.

### Communications nationales

Donfack Guefack, V., Bertaud-Gounot, V., Brillet, E., & Duvauferrier, R. (2011). **Création d'une interface utilisateur et d'une base de règles pour système expert à partir d'une ontologie.** Recherche en Imagerie et Technologie pour la santé (RITS), Rennes-France, Avril 2011.

Donfack Guefack, V., Bertaud-Gounot, V., Burgun, A., Barton, A., Duvauferrier, R., & Lasbleiz, J. (2012). **Ontologie sémiologique biomédicale et sémiologie quantitative.** IC pour l'Interopérabilité Sémantique dans les applications en e-Santé, Atelier associé à la conférence IC 2012 (pp. 1–6). Paris: IC'2012.

Donfack Guefack, V., Serrand-Obry, F., Lasbleiz, J., Duvauferrier, R., & Bertaud-Gounot, V.. (2012). **Système d'aide à la décision à base d'ontologie en urgence odontologique.** Biologie Santé et Innovation Technologique (BIOSIT) (Ed.), *Journee Doctorants Post-Doctorants* (p. 1). Rennes : Biologie Santé et Innovation Technologique (BIOSIT).

Serrand-Obry, F., Donfack Guefack, V., Bertaud-Gounot, V., Massot, M. & Limbour, P. (2013). **Développement d'une ontologie des urgences buccales : vers la conception de logiciel d'aide au diagnostic.** Innovations et Nouvelles Technologies Implantologie (p. 1). Lille : 60<sup>ème</sup> congrès de la Société Française de Chirurgie Orale (SFCO).

### Poster

Serrand-Obry, F., Donfack Guefack, V., Lasbleiz, J., Duvauferrier, R., & Bertaud-Gounot, V. (2012). **Elaboration d'une ontologie des urgences odontologiques : vers un système**

**d'aide à la décision.** Journées ouvertes Biologie Informatique Mathématique (JOBIM) (Ed.), Rennes.

## **En revue**

Donfack Guefack, V., Lasbleiz, J., Duvauferrier, R., Bourdé, A., Morelli, J., & Bertaud-Gounot, V. (2013). **Ontology driven decision support systems based on abductive and probabilistic reasoning: a case study of the knee injurie.** Artificial Intelligence in Medicine (AIIM).

VU :

VU :

**Le co-directeur de Thèse**

**Le Responsable de l'École Doctorale**

Valérie **BERTAUD-GOUNOT**

**VU pour autorisation de soutenance**

**Rennes, le**

**Le Président de l'Université de Rennes 1**

**Guy CATHELINEAU**

**VU après soutenance pour autorisation de publication :**

**Le Président de Jury,**